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

Lukas Altermatt† Hugo van Buggenum‡ Lukas Voellmy§

5th May 2022

Abstract

We develop a dynamic general equilibrium model of bank runs and solvency crises. Banks hold cash as a short-term asset and invest in government bonds and loans to entrepreneurs as long-term assets. Impatient households withdraw cash from the bank in order to purchase goods from entrepreneurs. If a run occurs, consumption by impatient households decreases, as some cash ends up with patient households. Therefore, the occurrence of a run lowers the profits of entrepreneurs and thus reduces the return on bank assets – the bank run triggers a solvency crisis through its effects on the real economy. This mechanism makes it individually rational to run the bank if other households do so, even if banks react to runs with deposit freezes. We also study whether runs can be prevented with combinations of deposit freezes, redemption penalties, and the provision of emergency liquidity by central banks.

Keywords: *Fragility, deposit freezes, emergency liquidity.*

JEL codes: E4, E5, G2.

*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 CO4 3SQ, United Kingdom. `lu.altermatt@gmail.com`

‡ETH Zürich, Zürichbergstrasse 18, 8092, Zürich, Switzerland. `hvanbuggenum@ethz.ch`

§Swiss National Bank. `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 believe 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. Further, since bank assets pay lower returns or default at higher rates in a recession, the economic downturn might cause a solvency crisis for banks that would have been perfectly solvent otherwise. In turn, this solvency crisis may rationalise the run in the first place. Thus, financial crises and recessions may arise endogenously and reinforce each other. A similar argument was made by Ben Bernanke in the wake of the 2007-2009 financial crisis:[2]

> *As we saw last fall [2008], (...) falling asset prices and the collapse of lender confidence may create financial contagion, even between firms without significant counterparty relationships. In such an environment, the line between insolvency and illiquidity may be quite blurry.*

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 illiquid assets held by banks is endogenous and varies with aggregate demand. Specifically, banks make loans to entrepreneurs who sell goods to households. 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. In turn, 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 illiquid 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 illiquid 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.[3]

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

[2] This quote is from a speech by Ben Bernanke at Jackson Hole in August 2009: `https://www.federalreserve.gov/newsevents/speech/bernanke20090821a.htm`

[3] 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.

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 – but 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 always stops runs, and it prevents them ex-ante if and only if the central bank purchases the assets at face value. If the central bank instead buys these assets at a discount, some demand shortfall and thus some default by entrepreneurs is inevitable, meaning that running the bank is still rational for patient depositors. In a sense, this result goes against Bagehot (1873), who advocated for central banks that lend to solvent, but illiquid banks at a high rate of interest. While doing so seems natural especially when the line between insolvency and illiquidity is blurry, it turns out that the liquidity problem turns into a solvency problem if and only if the central bank offers emergency liquidity at a penalty rate, i.e., buys assets below face value.

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 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. The provision of emergency liquidity at a discount can be interpreted as a fire sale in our model, but also in this case, the fundamental asset value falls because providing emergency liquidity at a discount does not prevent the bank run.

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 by impatient households, banks invest some of the deposits in illiquid assets, 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 found 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, so banks cannot offer liquidity insurance, but capital investment and therefore DM consumption is higher relative to 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.[4] 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. This provision of emergency liquidity always stops runs, and prevents them if and only if the government buys the assets at face value. In this case, banks can continue paying out the promised amount to impatient depositors, which stabilises aggregate demand, thereby prevents default by entrepreneurs, and thus gives patient depositors no reason to run in the first place. If the government buys these assets at a discount however, banks will run into a solvency crisis unless they reduce payouts to impatient depositors, which leads to a shortfall in demand and some default by entrepreneurs, and thus running the bank remains rational for patient depositors ex ante. 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.

[4] Deposit freezes are sometimes also referred to as suspension of convertibility in the literature.

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 labor as inputs) is used to produce goods sold against fiat money.[5,6] The model of Aruoba et al. (2011) is itself based on the New Monetarist framework (Kiyotaki and Wright (1989), Lagos and Wright (2005)). Similar 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.[7]

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.[8] 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

[5]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 the DM.

[6]Interestingly, the double holdup problem discussed in Aruoba et al. (2011) - high bargaining power for buyers implying sellers cannot recoup the cost of investment in capital; high bargaining power for sellers implying that buyers cannot recoup the cost of carrying real balances - is resolved in our paper. The reason is that our sellers (which we call entrepreneurs) finance capital investment through loans, and since these need to be repaid after the DM takes place, capital investment is not subject to a holdup problem.

[7]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.

[8]It is well known that full deposit freezes are generally not desirable in the presence of aggregate uncertainty, that is, 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.

limited commitment, deposit freezes are much less effective in preventing runs, even in the absence of aggregate uncertainty.[9] The present paper adds a third, distinct reason why deposit 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.

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.[10] 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, both of these papers study endowment economies, so they cannot speak to the mutually reinforcing effects of financial crises and real recessions. Also, they 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 helps depositors to insure against random liquidity needs in a way that is reminiscent of Berentsen et al. (2007) and of Williamson (2012).[11] Similar to our paper, Andolfatto et al. (2019) combine the Lagos-Wright and Diamond and Dybvig (1983) frameworks. They mostly focus on the coexistence of markets and banks, and how the presence of markets

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

[10]See also Liu (2019) and Goldstein et al. (2020) and Amador and Bianchi (2021) for models of systemic bank panics with fire sales.

[11]In Berentsen et al. (2007), banks intermediate cash between agents that do not want to consume and those with consumption needs. In our model, as in Williamson (2012), banks pool agents' cash and assets and issue demand deposits.

affects banks' ability to provide liquidity insurance, topics that are not the focus 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.

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.[12] 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 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

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

[12] 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.

where $c(q,k)$ represents the amount of labour required to produce q units of the consumption good given k.[13] 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. Further, we assume entrepreneurs also discount the next period by a factor β in the interest of symmetry with households, so the utility function of a young entrepreneur at the beginning of the CM is 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.[14] 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 this economy, 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 $1 + \pi_{t+1} \equiv \frac{\phi_t}{\phi_{t+1}}$.

[13] As an example, suppose $f(k,h) = k^\alpha h^{1-\alpha}$ (Cobb-Douglas), with $\alpha \in (0,1)$. Then, $c(q,k) = q^{\frac{1}{1-\alpha}} k^{-\frac{\alpha}{1-\alpha}}$.

[14] 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 $\ell_t = \dfrac{k_t}{\phi_t}$. 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.[15]

Besides 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 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:

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

[15] 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. In particular, 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.

$$s.t. \quad l_t = x_t + nx_t^e + nk_t$$
$$h_t = c(q_t^e, k_t)$$
$$\theta q_t = nq_t^e$$

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 simply the DM production function, saying that hours worked by entrepreneurs in the DM must be sufficient to produce desired output q_t^e given capital stock k_t. The third constraint says that consumption of the DM good by impatient households must equal production of the DM good by entrepreneurs. 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.[16,17]

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

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

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

Households' optimal CM consumption (x^*) is pinned down by condition (7), which says that households' marginal utility of consuming the CM good must equal the marginal disutility of producing it. 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. We will restrict attention to symmetric monetary equilibria, meaning that $\phi_t > 0$ for all t.[18]

[16] 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.

[17] 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 $c_k(q^e, \kappa) = c_k(1, \kappa)$.

[18] There is always an equilibrium with $\phi_t = 0$ for all 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.

Households. In the CM, households solve[19]

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

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), 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

$$U'(x_t) = 1 \tag{11}$$
$$u'(q_t) = \rho_t\left(1 + \frac{\iota_{t+1}}{\theta}\right) \tag{12}$$
$$\iota_{t+1} = i_{\ell,t+1} \tag{13}$$
$$\iota_{t+1} = i_{b,t+1}. \tag{14}$$

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:

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

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.

[19] 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 B.1.

The second constraint says that the capital stock must be financed with a corresponding nominal loan. The last condition simply follows from the DM production function. Inserting the constraints into the objective function, we can reformulate this problem as

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

Given the choice of k_t, we need to check whether entrepreneurs are willing to work in the DM, which is the case if

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

From this, it is clear that there is no partial default in equilibrium: If entrepreneurs find themselves in a situation where the maximised return from production is not enough to cover the cost of the loan and the disutility of working, they are better off not producing and working at all. Assuming the above constraint is slack, the first-order conditions to the entrepreneur's problem are

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

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

In Appendix B.2, we verify that the constraint is indeed slack given these FOCs.[20]

Equilibrium. First, note that there is no default in equilibrium, since under rational expectations all agents know how high 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

$$nk_t = \phi_t \ell_t, \tag{17}$$

while market clearing in the DM is

$$nq_t^e = \theta q_t. \tag{18}$$

Combining conditions (12), (13), (15) and (16), and using the fact that $\tilde{i}_{\ell,t+1} = i_{\ell,t+1}$, we get that the equilibrium levels of q_t and κ_t are determined by

$$\frac{u'(q_t)}{c_q(1, \kappa_t)} = 1 + \frac{\iota_{t+1}}{\theta} \tag{19}$$

$$-c_k(1, \kappa_t) = 1. \tag{20}$$

[20]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 $f(k, h)$.

It is easy to see that the Friedman rule ($\iota_{t+1} = 0$) implements the first-best allocation. Away from the Friedman rule, $\frac{\iota_{t+1}}{\theta}$ creates a wedge between the marginal utility of consuming the DM good and the marginal disutility of producing it; i.e., the higher inflation and the lower the probability of a household to turn out to be impatient, the further is q 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 (20) with (9) shows that κ is optimal in equilibrium. Put differently, 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.

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.[21] 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. In the following sections, we investigate whether the banking equilibria we find are prone to unexpected runs.

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.[22] Note, however, that this interest rate may differ from the Fisher rate ι_t, which is the opportunity cost of holding

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

[22] 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 in order to relax the constraints (21)-(25), but if $i_{\ell,t} \neq i_{b,t}$, they would strictly prefer to hold whichever asset pays the higher interest rate. However, if banks do not hold an asset, it must pay ι_t. Further, no asset can pay a higher interest rate than ι_t, as otherwise demand from households would be infinite. Thus, we can conclude that $i_{\ell,t} = i_{b,t} \; \forall t$ must hold.

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.[23] 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:

$$d_t^I = p_t q_t \tag{21}$$

$$m_t^b \geq \theta d_t^I \qquad (\beta\phi_{t+1}\zeta_t) \qquad \text{liquidity const.} \tag{22}$$

$$m_t^b + a_t^b(1+i_{t+1}) \geq \theta d_t^I + (1-\theta) d_{t+1}^P \qquad (\beta\phi_{t+1}\xi_t) \qquad \text{solvency const.} \tag{23}$$

$$d_{t+1}^P \geq d_t^I \qquad (\beta\phi_{t+1}(1-\theta)\psi_t)) \qquad \text{patient IC const.} \tag{24}$$

$$u(q_t) - u(0) \geq \beta\phi_{t+1} d_{t+1}^P, \qquad \text{impatient IC const.} \tag{25}$$

where the terms in brackets denote the Lagrange multipliers on the constraints. Equation (21) states that choosing d^I is equivalent to choosing the DM consumption of buyers, as banks take the DM price p_t as given. In contrast to individual households, banks can economise on cash holdings by exploiting the law of large numbers. Equation (22) is a liquidity constraint, saying that the bank must hold enough cash to make payments to early withdrawers. Equation (23) is a solvency constraint, saying that the total value of the banks' assets must be at least as high as the total promises it makes to depositors. Equation (24) is an incentive compatibility constraint for patient depositors, as the bank needs to make sure these depositors prefer to wait until the next CM to make their withdrawals. Finally, equation (25) 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 get from withdrawing in the CM. We will ignore this constraint for now and check later whether or not it binds.[24]

Taking the first-order conditions to the above problem and rearranging yields

$$\zeta_t = \frac{i_{t+1}(1+\iota_{t+1})}{1+i_{t+1}}, \quad \xi_t = \frac{1+\iota_{t+1}}{1+i_{t+1}}, \quad \text{and} \quad \psi_t = \frac{\iota_{t+1} - i_{t+1}}{1+i_{t+1}}, \tag{26}$$

i.e., the liquidity constraint is binding unless the nominal rate is zero, the solvency constraint is always binding (which just implies that banks make zero profits) and patient

[23]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.

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

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 satisfy patient depositors' IC constraint entails no opportunity cost for banks. Further, consumption in the DM is characterised by

$$u'(q_t) = \rho_t \left[1 + \frac{\theta \iota_{t+1}(1+i_{t+1}) + (1-\theta)(\iota_{t+1} - i_{t+1})}{\theta(1+i_{t+1})}\right], \tag{27}$$

and the asset demand schedule of the bank is

$$m_t^b \geq \theta p_t q_t \qquad \text{with equality if } i_{t+1} > 0 \tag{28}$$

$$m_t^b + a_t^b(1+i_{t+1}) \geq p_t q_t \qquad \text{with equality if } i_{t+1} < \iota_{t+1}. \tag{29}$$

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

$$nk_t = \phi_t(\ell_t + \ell_t^b), \tag{30}$$

and market clearing in the bond market is

$$B_t = b_t + b_t^b. \tag{31}$$

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

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

In a monetary steady state we have $\phi_t > 0 \ \forall t$, all real variables are constant, and the Fisher rate is $1 + \iota = \frac{1+\mu}{\beta}$, where μ is the money growth rate. Combining (15), (16) and (27), 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

$$\frac{u'(q)}{c_q(1,\kappa)} = 1 + \frac{\theta\iota(1+i) + (1-\theta)(\iota - i)}{\theta(1+i)} \tag{32}$$

$$-c_k(1,\kappa) = \frac{1+i}{1+\iota}. \tag{33}$$

Taking nominal rates i and ι as given, this system has a unique solution. Condition (33) shows that the capital stock per produced DM good, κ, is uniquely determined through i and ι. Comparing (33) 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

$$\rho = c_q(1,\kappa) = c(1,\kappa) - \kappa c_k(1,\kappa) \tag{34}$$

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 (32) and (34), we can see that changes in i have two opposing effects on 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 (32) 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

$$\mathcal{A} \equiv \phi_t a_t = nk + \mathcal{B} = \kappa q \theta + \mathcal{B}. \tag{35}$$

Combining aggregate asset supply (35) with the bank's asset demand schedule (28)-(29), we get that the asset market clears if and only if:[25]

$$\mathcal{A} \begin{cases} \geq (1-\theta)\rho q & \text{if} \quad i = \iota \\ = \frac{1+\iota}{1+i}(1-\theta)\rho q & \text{if} \quad i \in (0,\iota) \\ \leq (1+\iota)(1-\theta)\rho q & \text{if} \quad i = 0 \end{cases} . \tag{36}$$

Definition 1. *A stationary monetary equilibrium (SME) is given by* $(q, \kappa, \rho, i, \mathcal{A})$ *that satisfy* (32)-(36).

The existence proof for SME is given in Appendix B.3:

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 $i \in (0, \iota)$. Condition (36) shows that, roughly speaking, 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. DM production satisfies $\frac{u'(q)}{\rho} = 1 + \iota$, meaning that consumption and investment levels are the same as in an economy without banks and with $\theta = 1$, that is, an economy without liquidity risk. Banks are saturated with assets in an FLI equilibrium, so that asset prices fall to the point where assets pay the Fisher rate. Large $\mathcal{B}$ increases the supply of illiquid assets and thus makes the existence of an FLI equilibrium more likely. 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.

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

In a ZLB equilibrium, we have $\kappa > \kappa^*$, i.e., capital per produced DM good is higher than optimal. Equation (32) reduces to $\frac{u'(q)}{\rho} = 1 + \frac{\iota}{\theta}$, which is the same condition as in the economy without banks. However, since κ is higher 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, in a ZLB equilibrium compared to the no-bank equilibrium.[26] 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.[27]

Let us now return to the question whether the IC constraint for impatient depositors (equation (25)) 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

$$\phi a^b \leq \bar{\mathcal{A}} \equiv (1-\theta)u(q), \tag{37}$$

that is, 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 total supply of illiquid assets given in (36) to be below the upper bound on the banks' holdings of illiquid assets given in (37), which is true 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 $\mathcal{A} > \bar{\mathcal{A}}$.

Coexistence of equilibrium cases. While the main focus of this paper is on unexpected bank runs, we want to briefly highlight some aspects of steady-state equilibria here. To

[26]Whether welfare is higher in a ZLB equilibrium compared to 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 labor.

[27]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, lower q also reduces demand for illiquid assets since it becomes easier for banks to satisfy patient depositors' IC, as a result of lower payouts to impatient depositors. It can be shown that the second 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.

do so, we assume that the DM production function is Cobb-Douglas:

$$f(k,h) = k^{\alpha} h^{1-\alpha}. \tag{38}$$

Appendix A discusses the properties of all three equilibrium types given this parametrisation and how to derive the results presented in this section in detail. Here, we only want to highlight the main findings.

Proposition 2. *If the DM production function is Cobb-Douglas and $\mathcal{B} = 0$, FLI equilibria exist for $\theta \geq \frac{1}{1+\alpha}$, while ZLB equilibria exist for $\theta \leq \frac{1}{1+\alpha}$. For $\theta = \frac{1}{1+\alpha}$, all three equilibrium cases coexist, with q varying across equilibrium cases.*

This shows that in an economy without government bonds, θ and α fully determine the equilibrium case the economy ends up in. For a given α, the economy will be in an FLI equilibrium for high values of θ and in a ZLB equilibrium for low values of θ. For fixed θ, an FLI equilibrium becomes more likely if α is high, which is natural given that α denotes the share of DM revenue that goes to capital (i.e., to banks in our model). For the knife-edge case $\theta = \frac{1}{1+\alpha}$, all three equilibria coexist, and any $i \in [0, \iota]$ constitutes an equilibrium.[28] Importantly, this multiplicity also has real effects. Figure 1 shows DM consumption over the domain of θ, with the dashed line depicting ZLB equilibria, the dotted line depicting FLI equilibria, and the solid line depicting PLI equilibria. These results show that there is a nonmonotonicity in q over θ, and further that when all equilibria coexist, the equilibria vary in terms of allocations and welfare.[29] As described above, an increase in i has two opposing effects on DM activity: a negative effect by increasing the DM price ρ and a positive effect by lowering the effective cost of liquidity. It turns out that in an economy with a Cobb-Douglas production technology and $\mathcal{B} = 0$, the first effect dominates at the level of θ where we have equilibrium multiplicity.

Proposition 3. *If the DM production function is Cobb-Douglas and $\mathcal{B} \neq 0$, all three equilibrium cases may coexist for a set of parameters with positive mass.*

When loans to entrepreneurs are not the only illiquid asset banks can invest in, the multiplicity of equilibria is not eliminated. Instead, it expands from only a knife-edge case to a parameter space with positive mass, making the multiplicity even more relevant. While we think this finding is interesting and could be important for our understanding of zero-lower bound episodes, it is not the focus of the remainder of this paper. Instead, all further results presented do not hinge on the equilibrium multiplicity and thus also not on the functional form assumed for $f(k,h)$ in equation (38).

[28] As we show in Appendix B.3, this is an instance of a general property of the model: whenever both an FLI and a ZLB exist, then (at least one) PLI equilibrium exists as well.

[29] Whether welfare is higher in a ZLB or an FLI equilibrium when they coexist is unclear. DM output in the ZLB equilibrium is higher, which may move DM output closer to its efficient level. However, a given DM output is produced with an inefficiently high amount of capital in a ZLB equilibrium, while the capital-labor mix is efficient in an FLI equilibrium.

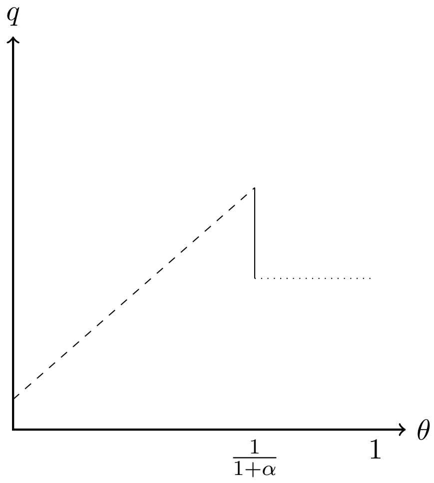

Figure 1: DM consumption as a function of θ in an economy with $\mathcal{B} = 0$.

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 steady-state values of variables. Following Diamond and Dybvig (1983) and much of the banking literature we assume sequential service, that is, 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 higher payout compared to not withdrawing.[30] 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 on aggregate compared to 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

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

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 equal 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, i.e., as long as policymakers do not step in when a run occurs, as we assume in Section 7. In this case, a run does not affect the real price of the DM good and we get a particularly simple expression for the share of active entrepreneurs:

Proposition 4. *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 ($\rho = \rho_S$)

(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, so 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 to an increase in the aggregate supply of DM goods while aggregate demand for the DM good decreases. Finally, given $\rho = \rho_S$ and $\pi = \pi_S$, item (ii) in Proposition 4 follows immediately from (39). ■

For banks, the effective gross nominal return on loans in a run equals $\chi(1 + i_S)$.[31] Notice that the return to an individual bank's portfolio depends on the redemption behaviour of

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

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 4 is the main innovation in our model relative to the existing literature.

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 changes in the capital share caused by the run:

Proposition 5. *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 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 its loan if and only if

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

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

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

Combining (39) and (43) then leads to the result in Proposition 5 ■

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, which do not rely on a public backstop such as emergency liquidity, a lender of last resort, or deposit insurance. In line with the bulk of the bank run literature, we assume the banking system reacts as a single, consolidated entity to runs.[32] The speed at which the queue is served is the same at all banks, that is, 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 emergency liquidity in the form of a troubled-asset relief program (TARP), i.e., the government stands ready to purchase illiquid assets from the banks, possibly at a discount. In that setting, we also allow banks to impose redemption penalties and deposit freezes once they observe a run is underway.

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 sets a TARP in place.[33] 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. 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.[34] 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 sets up a TARP after observing a run (off the equilibrium path), we will say that these measures *prevent runs*.

[32] 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, different to the banking authority in Ennis and Keister (2009, 2010), the consolidated banking system does not suffer from limited commitment in our model.

[33] We do not model explicitly how banks realise that a run is underway. One might imagine that banks observe (with some lag) that the rate of redemptions is higher than usual.

[34] Note that in the absence of a TARP, since assets cannot be liquidated prematurely, banks will impose a full deposit freeze by default once they run out of cash in the DM.

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 $\lambda = \theta$ and $\eta = 0$.

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.[35]

In addition to partially freezing deposits, banks may impose a penalty (or haircut) on early redemptions, which we denote by σ. Depositors who redeem 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.[36,37]

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

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

[35] 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 $\frac{\partial \overline{\overline{\eta}}}{\partial \lambda} \leq 0$.

[36] An equivalent formulation would be to say that banks pay some amount (d^{DM}, d^{CM}) to depositors redeeming 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 redeemed in the DM equally among the depositors who did not redeem in the DM.

[37] The term 'partial suspension' has sometimes been used in the literature to describe the case with $\sigma = 1-\eta$. We find it useful to explicitly distinguish between deposit freezes, where funds are 'frozen' but not lost, and penalties on redemptions.

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 equals left-over cash not paid out in the DM and the second term equals the proceeds from a bank's asset portfolio. 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 settings 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 $d_S^I > d_R^P(\cdot)$.

Next, 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 4, we then have

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

for $\eta \geq \overline{\eta}$.[38] Combining (47) and (48), we get that the banking system is fragile under any deposit freeze satisfying $\eta \geq \overline{\eta}$ iff

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

[38] 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 $(1-\chi)$, setting $\eta < \overline{\eta}$ unambiguously exacerbates incentives to run compared to a freeze with $\eta \geq \overline{\eta}$, allowing us to focus on $\eta \geq \overline{\eta}$ without loss.

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 6. *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 with $\mathcal{A} < \min\{\bar{\mathcal{A}}, \bar{\mathcal{A}}^F\}$, 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 $\rho \equiv \dfrac{\phi}{1+\iota} p$, $\phi \ell^b = \theta q \kappa$, $\mathcal{A} \equiv \phi(b^b + \ell^b)$ 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 $\mathcal{A} \leq \bar{\mathcal{A}}$), they are fragile iff $\mathcal{A} \leq \bar{\mathcal{A}}^F$.[39] ■

The fragility of the banking system in ZLB and PLI equilibria is closely connected 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 steady state, so 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. The effect of θ can be seen nicely when considering the case with $\mathcal{B} = 0$ and $f(k, h)$ given by (38), i.e., Cobb-Douglas. Then, the existence condition for FLI equilibria from (36) together with the result in proposition 6 imply that a run-prone FLI equilibrium exists iff $\frac{1}{\alpha} \in \left[\frac{\theta^2}{1-\theta}, \frac{\theta}{1-\theta}\right)$.

A remarkable result of proposition 6 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

[39] 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, the less (more) likely it is that banks are fragile. To abstract from this, we assume for the remainder of the paper that either $\mathcal{A} \leq \bar{\mathcal{A}}$ or $\bar{\mathcal{A}}^F \leq \bar{\mathcal{A}}$ holds, such that the relevant condition for FLI equilibria to be fragile is $\mathcal{A} \leq \bar{\mathcal{A}}^F$.

after a run has started, the fact that impatient depositors can withdraw less cash due to the freeze makes running optimal for patient depositors. This latter point is the major difference to 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 however, 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.

How should banks set η in order to minimize ex post welfare losses caused by a run, given that they cannot prevent runs with pure deposit freezes? 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 7. *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 $\eta \in [\overline{\eta}, \overline{\overline{\eta}}]$.[40]

[40]Footnote 38 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

A given mix of deposit freezes and penalties on redemptions, (η, σ), 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 in the DM even if (hypothetically) all other depositors continue running after banks impose (η, σ).[41] If all depositors continue to redeem in the DM after banks impose (η, σ), we have $\omega = \overline{\omega}(\eta)$ and $\chi = m_S^b / d_S^I$ (for the latter, see the discussion in the previous subsection). From (47), we get 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, (η, σ), 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 8. *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 B.4.1 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, (η, σ), *prevents* a run if the prospect of banks imposing (η, σ) after realizing that a run is underway eliminates patient depositors incentive 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 $\omega = \theta(1 - \lambda)$ and $M^{RE} = \theta\left[\lambda + (1 - \lambda)\eta\right] d_S^I$. By Proposition 4, the share of non-defaulting entrepreneurs then equals

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

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

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

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 the available cash is distributed more evenly among impatient depositors in the event of a run. See also Footnote 43 below.

[41] 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.

Of course, impatient depositors must be willing to redeem in the DM (and incur the redemption penalty) rather than waiting until the CM, which requires[42]

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

In Appendix B.4.2, 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

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

In Appendix B.4.3, we show that the condition in (55) can be rewritten as a lower bound on the redemption penalty, $\sigma \geq \underline{\hat{\sigma}}(\eta)$. 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).[43]

Proposition 9. *Even with $\lambda = 0$, there may be no policy (η,σ) that prevents a run.*

We prove this proposition by providing an example in Appendix B.6 where there is no (η,σ) preventing runs even if $\lambda = 0$. However, in many cases banks can prevent runs by setting (η,σ) appropriately, provided that λ is low enough; 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

[42] 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.

[43] In conditions (55) and (54), we have implicitly required that banks' reaction to runs, besides stopping the run, be such that all impatient depositors are able and willing to redeem in the DM. This is without loss of generality, in the sense that 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.

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' incentive 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), 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, 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 effectively 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 remain to withdraw. We denote this level with $\eta^{\max}$.

Proposition 10. *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 B.5 for the proof and the derivations related to Proposition 10. 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 $\eta^{\max} = \hat{\eta}^{max}$).

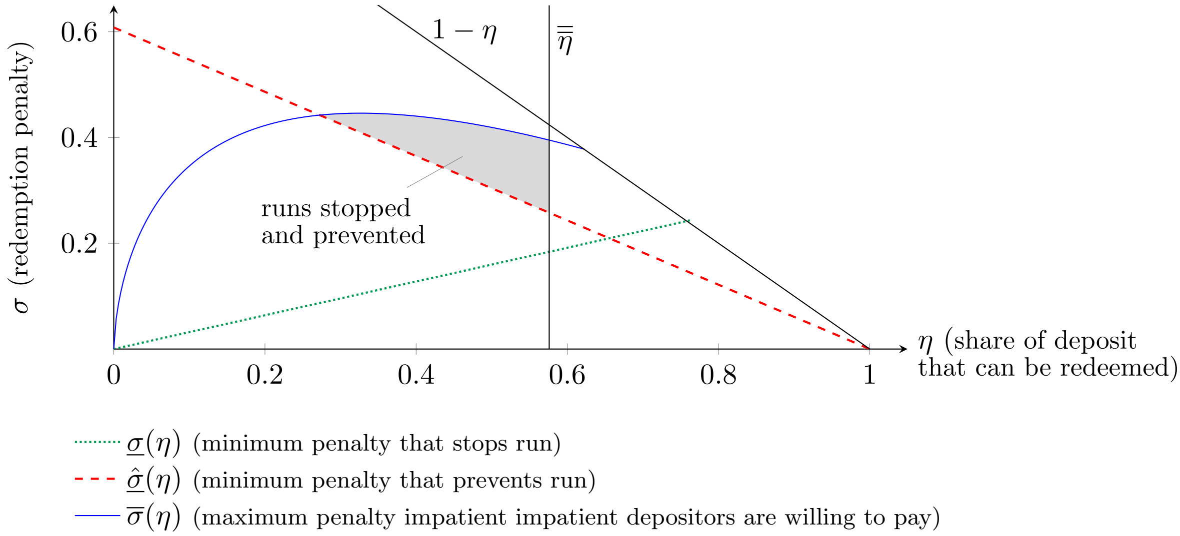

Figure 2: Example where runs can be prevented.

Example where Runs can be Prevented

Figure 2 provides a numerical example illustrating how (η, σ) needs be set in order to both stop and prevent runs. We assume a Cobb-Douglas production technology as specified in (38), and $u(q) = q^{\nu}$. The parameters used for this example are summarized in Table 1. Since $\mathcal{B} = 0$ and $\theta < \frac{1}{1+\alpha}$, this economy is in a zero lower bound steady state equilibrium

Table 1: Parameter values for Figure 2.

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

(see Section 5). The grey area depicts the set of (σ, η) that both stop and prevent runs. The fact that $\underline{\sigma}(\eta)$ is strictly increasing in η while $\hat{\underline{\sigma}}(\eta)$ is strictly decreasing in η is not specific to the example in Figure 2. As we show in Appendix B.4, this is always the case if banks are fragile under pure deposit freezes. Intuitively, higher DM payouts increase the incentive to run, which requires a higher redemption penalty in order to stop patient depositors from running. On the other hand, 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' incentive to participate in the run in the first place and lowering the minimum redemption penalty necessary to deter them from running. This is a characteristic property of our model – increasing payouts to impatient depositors can lower patient depositors' incentive to run due to the tight connection between the banking systems' liabilities and assets.

8 Emergency Liquidity

We now turn to the role of policy in stopping or preventing runs. We focus on a policy intervention which provides emergency liquidity to the banking system by means of a troubled-asset relief program (TARP). This means that banks can obtain liquidity from the government by selling their (troubled) illiquid assets in exchange for money.

We assume that the government cannot increase its real indebtedness, which restricts the ability to provide liquidity in real terms.[44] Specifically, we have

$$\frac{M + (1 + i_S)B}{M_S + (1 + i_S)B_S} \leq \frac{1 + \pi}{1 + \pi_S}, \tag{56}$$

where $M + (1 + i_S)B$ denotes the governments' 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 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 thus bounded from above. 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 in the TARP mature.

To provide emergency liquidity in case of a run, the government stands ready to convert 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 haircuts or 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 $\delta_\ell \leq \delta_b$, as bonds are safer assets compared to 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

$$m_S^b - \theta d_S^I + \delta_b(1 + i_S)b_S^b + \delta_\ell(1 + i_S)\ell_S^b. \tag{57}$$

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

[44] This constraint can be motivated in various ways. If TARP is performed by the central bank, the constraint may capture the central bank's inability to raise taxes. If TARP is performed by the fiscal authority instead, it may capture political constraints, such as a debt limit.

liquidity:[45]

$$\eta = 1 - \sigma = \min\left\{ \frac{m_S^b - \theta d_S^I + \delta_b(1+i_S)b_S^b + \delta_\ell(1+i_S)\ell_S^b}{(1-\theta)d_S^I}, 1 \right\}. \tag{58}$$

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

$$M^{RE} = \theta\left[\theta + \eta(1-\theta)\right] d_S^I. \tag{59}$$

8.1 Inflation and Defaults with Emergency Liquidity

We start with the observation that – according to proposition 5 – changes in inflation have no direct effect on entrepreneurs' incentive to produce in the DM and repay loans. Intuitively, 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.

Next, let us consider the implications of the government's intervention for inflation. These depend on the type and quantity of assets purchased by the central bank. Let τ_b and τ_ℓ denote the fraction of bonds and loans that banks liquidate, 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 has withdrawn. In case $m_S^b - \theta d_S^I < \eta(1-\theta)[\theta + \mathcal{I}(1-\theta)]d_S^I$, the bank needs to sell assets in order to meet withdrawals, and the asset sales are such that

$$\eta(1-\theta)\left[\theta + \mathcal{I}(1-\theta)\right] d_S^I - (m_S^b - \theta d_S^I) = \tau_b\delta_b(1+i_S)b_S^b + \tau_\ell\delta_\ell(1+i_S)\ell_S^b. \tag{60}$$

The LHS in equation (60) 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 central bank. Nominal liabilities of the government at the beginning of the next CM, just after the active entrepreneurs have repaid their loans, then become

$$d_S^I\{\theta + \eta(1-\theta)[\theta + \mathcal{I}(1-\theta)]\} - \tau_\ell\chi(1+i_S)\ell_S^b + (1-\tau_b)(1+i_S)b_S^b + (1+i_S)(B_S - b_S^b), \tag{61}$$

where we use $M_S = m_S^b$. The first term in (61) is the money in circulation after the injection of emergency liquidity, the second term is the money withdrawn from circulation

[45] Given that banks impose redemption penalty (58) when realising a run is underway, the remaining cash is distributed equally across the remaining measure $1 - \theta$ of depositors in case the run continues. If banks choose a lower redemption penalty, then a run cannot be stopped once it has started since the depositors who show up last receive nothing. If the run is not stopped, impatient depositors will end up with less cash than when the penalty is set according to (58), which implies worse macroeconomic outcomes.

in the beginning of the CM through loan repayments by entrepreneurs, and the third and fourth terms are government bonds held by banks and households respectively. Combining (60) with (61), we can express nominal government liabilities after entrepreneurs have repaid loans as

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

Combining (62) with (56), we obtain[46]

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

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 by the government. Put differently, emergency liquidity can only be inflationary if the government makes losses on its asset purchases. As long as the government does not make losses with its intervention, the amount of money withdrawn from circulation when assets are repaid is at least a high as the amount injected in the run – the provision of emergency liquidity does therefore 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 11. *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 11 can be found in Appendix B.7. The intuition is as follows. Since inflation has no direct effect on defaults, $\chi < \delta_\ell$ can only occur if there is a decline in the capital share α, as defined in (41). However, there is no equilibrium consistent with a decrease in the capital share. The reason for this last point is that a decline in the capital share α has two effects on entrepreneurs. On the one hand, it generates inflation through losses on the government's liquidity provision. This reduces the real debt burden for the entrepreneurs. On the other hand, a decline in α also reduces the share of real revenues that are left after compensating entrepreneurs for their disutility of production. It turns out that the latter effect always dominates; if α declines, the entrepreneurs' real repayment burden reduces by less than the revenue left to repay loans, implying that entrepreneurs are better off defaulting. Since this is true for all entrepreneurs, supply collapses, but since aggregate demand is still positive, this is not consistent with equilibrium. Thus, we can rule out a case where the provision of emergency liquidity leads to a decline in the capital share, and this in turn rules out inflation caused by emergency liquidity provision.

[46] 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.

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

Proposition 12. *$\pi = \pi_S$ also holds in case of a run if entrepreneurs' gross repayment is fixed in real terms.*

The proof for Proposition 12 is in Appendix B.8. 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: One 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, the two forces described above cancel against each other. Thus, the effects of emergency liquidity provision are the same with nominal and real debt contracts.

Finally, we get from (59) and our results from Propositions 4 and 11 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 $\eta < 1$, i.e., if and only if banks charge a penalty on DM redemptions. According to (58), whether or not banks charge a redemption penalty depends on the discounts (δ_ℓ, δ_b) which the government imposes on its asset purchases. In particular, we have

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

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 $d_S^I \leq m_S^b + (1+i_S)(\ell_S^b + b_S^b)$ 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 to default on loans by a fraction of the entrepreneurs. 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. Despite the fact that the government cannot create liquidity in real terms by simply increasing lump-sum taxation, by buying illiquid assets at prices reflecting nominal face value, a run can thus be prevented from spilling over to the real economy.

8.2 Preventing Runs with Emergency Liquidity

Let us now consider whether a run stops after it has been detected and the redemption penalty is imposed, and whether runs can be prevented by the prospect of a TARP. 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 exactly 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 (58) yields the following result:

Proposition 13. *Runs can always be stopped with the provision of emergency liquidity, independent of the discounts imposed on asset purchases.*

Proof. Suppose first 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 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 11). ■

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 not only ensures that the government does not make losses on its asset purchases, it also ensures 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 14. *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, emergency liquidity prevents runs if and only if the discounts are small and/or banks hold many illiquid assets.*

Proof. Note that condition (67) can be rewritten as $d_S^I \leq m_S^b + (1+i_S)(b_S^b + \chi \ell_S^b)$. Consider first ZLB and PLI equilibria. As seen earlier, the IC constraint of patient depositors binds in these equilibria, so that we have $d_S^I = d_S^P = m_S^b + (1+i_S)(\ell_S^b + b_S^b)$. Condition (67) 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. In these equilibria, the IC constraint of patient depositors is slack in steady state, implying $d_S^I \leq m_S^b + (1+i_S)(\ell_S^b + b_S^b)$, with strict inequality except for a knife-edge case. We thus get that banks are only fragile if $\chi < \chi' \leq 1$, where χ' is some critical threshold. From Corollary 1 we know that, in FLI equilibria, χ is decreasing in the discounts and increasing in the difference between steady state payouts to patient and impatient depositors. ■

Intuitively, if discounts are imposed, some impatient depositors may end up with less cash compared to 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 which impatient households want to consume. We showed that introducing 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 also able to invest in government bonds besides 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 troubled assets, runs can be prevented if and only if the government is willing to purchase these assets at face value. 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 illiquid assets, so running the bank remains rational for patient depositors.

Bibliography

Amador, M. and Bianchi, J. (2021). Bank runs, fragility, and credit easing. *Federal Reserve Bank of Minneapolis Working Paper 785.*

Andolfatto, D., Berentsen, A., and Martin, F. (2019). Money, banking and financial markets. *Review of Economic Studies*, page forthcoming.

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.*

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.

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.

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

Keister, T. and Mitkov, Y. (2019). Bailouts, bail-ins and banking crises. Crc tr 224 discussion paper series.

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.*

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

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.

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

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 Banking Equilibrium for a Cobb-Douglas Production Function

To derive the results presented in Propositions 2 and 3, in this section we consider matters when the DM production function is Cobb-Douglas, i.e.,

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

with $\alpha \in (0,1)$. For this parameterization, we can rewrite equation (33) to get

$$\kappa = \left(\frac{\alpha}{1-\alpha} \frac{1+\iota}{1+i} \right)^{1-\alpha}, \tag{68}$$

and thus the real DM price from equation (34) is

$$\rho = \frac{1}{1-\alpha} \left(\frac{1-\alpha}{\alpha} \frac{1+i}{1+\iota} \right)^{\alpha}. \tag{69}$$

From equation (32) we get that DM output in an FLI and a ZLB equilibrium respectively equals

$$q^{FLI}(\alpha,\theta,\iota) = u'^{-1} \left(\frac{1+\iota}{1-\alpha} \left(\frac{1-\alpha}{\alpha} \right)^{\alpha} \right), \quad q^{ZLB}(\alpha,\theta,\iota) = u'^{-1} \left(\frac{1+\iota/\theta}{1-\alpha} \left(\frac{1-\alpha}{\alpha(1+\iota)} \right)^{\alpha} \right) \tag{70}$$

and for PLI equilibria we get:

$$q^{PLI}(i;\alpha,\theta,\iota) = u'^{-1} \left(\frac{1}{1-\alpha} \left(1 + \frac{\theta\iota(1+i) + (1-\theta)(\iota - i)}{\theta(1+i)} \right) \left(\frac{1-\alpha}{\alpha} \frac{1+i}{1+\iota} \right)^{\alpha} \right). \tag{71}$$

Given this, we can now revisit the existence conditions (36) for the three equilibria. For the full liquidity-insurance equilibrium, the existence condition becomes

$$\frac{\mathcal{B}}{q^{FLI}(\alpha,\theta,\iota)} + \theta \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \geq \frac{1-\theta}{1-\alpha} \left(\frac{1-\alpha}{\alpha} \right)^{\alpha}, \tag{72}$$

while for the zero-lower bound equilibrium it is

$$\frac{\mathcal{B}}{q^{ZLB}(\alpha,\theta,\iota)} + \theta \left(\frac{\alpha(1+\iota)}{1-\alpha} \right)^{1-\alpha} \leq (1+\iota) \frac{1-\theta}{1-\alpha} \left(\frac{1-\alpha}{\alpha(1+\iota)} \right)^{\alpha}, \tag{73}$$

and for the partial liquidity-insurance equilibrium, the existence condition is

$$\frac{\mathcal{B}}{q^{PLI}(i;\alpha,\theta,\iota)} + \theta \left(\frac{\alpha}{1-\alpha} \frac{1+\iota}{1+i} \right)^{1-\alpha} = \left(\frac{1+\iota}{1+i} \right)^{1-\alpha} \frac{1-\theta}{1-\alpha} \left(\frac{1-\alpha}{\alpha} \right)^{\alpha}, \tag{74}$$

with $i \in [0, \iota]$.

Equilibrium without Government Bonds

We consider first the specific case in which $\mathcal{B} = 0$. This implies that the only illiquid assets banks can invest in are loans to entrepreneurs. While this is not the most realistic and also not our preferred specification, it is very illustrative since it isolates the new mechanism we introduce here: The return on illiquid assets is not independent of the real economy, and if a bank run occurs, conditions in the real economy and thus returns on illiquid assets worsen.

With $\mathcal{B} = 0$, the existence conditions (72)-(73) reduce to

$$\theta \begin{cases} \geq \frac{1}{1+\alpha} & \text{Full liquidity-insurance equilibrium} \\ = \frac{1}{1+\alpha} & \text{Partial liquidity-insurance equilibrium} \cdot \\ \leq \frac{1}{1+\alpha} & \text{Zero-lower bound equilibrium.} \end{cases} \tag{75}$$

What do these results imply for DM consumption? First, note that q^{FLI} is independent of θ, while q^{ZLB} increases in θ. Third, it is also relatively straightforward to show that $q^{FLI}(\alpha, \theta, \iota) > \lim_{\theta \to 0} q^{ZLB}(\alpha, \theta, \iota)$. Thus, to characterise the path of q over the whole domain of θ, all that is left to determine is whether the q's differ in the three equilibria when $\theta = \frac{1}{1+\alpha}$, and if they do, which one delivers the highest q. Computing these two quantities shows that

$$q^{FLI}(\alpha, \theta, \iota) < \lim_{\theta \to 1/(1+\alpha)} q^{ZLB}(\alpha, \theta, \iota) \;\Leftrightarrow\; 1 + \iota > (1 + \iota(1-\alpha))\,(1+\iota)^{-\alpha},$$

which happens to hold for any $\iota > 0$ and $\alpha > 0$.

Equilibrium with Government Bonds

Having discussed matters without government bonds, we now turn towards the more interesting case with a positive amount of government bonds available in the economy. We can rewrite the existence condition for an FLI equilibrium (72) as:

$$\mathcal{B} \geq q^{FLI}(\alpha, \theta, \iota) \left(\frac{1-\theta}{1-\alpha} \left(\frac{1-\alpha}{\alpha} \right)^{\alpha} - \theta \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \right). \tag{76}$$

It is easy to see that for any ι and θ, there is a $\mathcal{B}$ that satisfies condition (76), thus showing that policymakers can always get the economy into an FLI equilibrium by supplying enough bonds. Rewriting the condition for a ZLB equilibrium, (73), in a similar fashion yields

$$\mathcal{B} \leq q^{ZLB}(\alpha, \theta, \iota) \left(\frac{1-\theta}{1-\alpha} \left(\frac{1-\alpha}{\alpha} \right)^{\alpha} - \theta \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \right) (1+\iota)^{1-\alpha}. \tag{77}$$

It is again straightforward that higher $\mathcal{B}$ makes it harder for the above condition to be satisfied, showing that by supplying enough bonds, policymakers can not only ensure that an FLI equilibrium exists, but also that a ZLB equilibrium does not exist. From the case with $\mathcal{B} = 0$, we know that an FLI and a ZLB equilibrium both exist when $\theta = \frac{1}{1+\alpha}$. An important question is thus whether there multiplicity of equilibria can still occur with $\mathcal{B} > 0$. This is the case whenever there exists a $\mathcal{B}$ that satisfies both conditions (76) and (77). A sufficient condition for this to be the case is

$$q^{ZLB}(\alpha, \theta, \iota) \geq \frac{q^{FLI}(\alpha, \theta, \iota)}{(1+\iota)^{1-\alpha}}. \tag{78}$$

To ensure that a $\mathcal{B}$ satisfying both (76) and (77) is positive, we also need

$$\theta < \frac{1}{1+\alpha},$$

which is of course a familiar expression from the previous analysis. We already showed that $q^{ZLB}(\cdot) > q^{FLI}(\cdot)$ when $\theta = \frac{1}{1+\alpha}$, and since $(1+\iota)^{1-\alpha} > 1$ for any $\iota > 0$, condition (78) is satisfied at this value of θ.[47] By continuity, the condition will also be satisfied by values of θ close to, but different from $\frac{1}{1+\alpha}$. Thus, there exist values of $\theta < \frac{1}{1+\alpha}$ ($\theta > \frac{1}{1+\alpha}$) for which the two equilibria coexist with a positive (negative) supply of government bonds. Further, as we show in Appendix B.3, coexistence of an FLI and a ZLB equilibrium implies existence of a PLI equilibrium. This shows that with $\mathcal{B} \neq 0$, all three types of equilibria coexist not only in a knife-edge case, but for a set of parameters with positive mass.

Appendix B Additional Derivations and Proofs

B.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.[48] 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,[49] we have:

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

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:

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

[47] At the Friedman rule, the FLI and ZLB equilibria are equivalent, so the condition always holds at equality.

[48] 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.

[49] 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.

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.

B.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

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

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.

B.3 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 (36) is fulfilled. Note that $\mathcal{A}, \rho$ and q can all be expressed as continuous functions of i. From (36), we get that an equilibrium with $i \in (0, \iota)$ exists iff:

$$Q(i) = 1 + i - \frac{(1-\theta)(1+\iota)\rho(i)q(i)}{\mathcal{A}(i)} = 0. \tag{79}$$

We also get from (36) 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 $i \in (0, \iota)$.

B.4 Derivations of the Thresholds for the Penalty on Early Redemptions

B.4.1 Threshold to Stop a Run

Rewriting the condition in (51) gives

$$\eta d_S^I \leq (\eta + \sigma) \underline{d}_R^P(\eta, \sigma). \tag{80}$$

Since the right-hand side of (80) 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 (80) is fulfilled iff $\sigma \geq \underline{\sigma}(\eta)$. Substituting for $\underline{d}_R^P(\eta,\sigma)$ we can rewrite condition (80) as:

$$\eta d_S^I \leq (\eta+\sigma)\frac{(1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right)}{1-\lambda-\overline{\omega}(\eta)(\eta+\sigma)}$$

$$\Leftrightarrow \sigma\left[(1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right) + \overline{\omega}(\eta)\eta d_S^I\right] \geq (1-\lambda)\eta d_S^I - \eta\left[(1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right) + \overline{\omega}(\eta)\eta d_S^I\right] \tag{81}$$

$$\Leftrightarrow \sigma \geq \left[\frac{d_S^I - m_S^b - (1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right)}{m_S^b - \lambda d_S^I + (1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right)}\right]\eta \equiv \underline{\sigma}(\eta)$$

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

$$\tilde{\eta}^{max} = \frac{(1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right) + m_S^b - \lambda d_S^I}{(1-\lambda)d_S^I} = \overline{\eta} + \frac{(1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right)}{(1-\lambda)d_S^I} \in (\overline{\eta}, 1), \tag{82}$$

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.

B.4.2 Threshold to Satisfy Impatient Depositors' Incentive Constraint

Rewriting the condition in (54) gives

$$u(\eta q_S) \geq \frac{\phi}{1+\iota}(\eta+\sigma)\overline{d}_R^P(\eta,\sigma). \tag{83}$$

Since the right-hand side of (83) 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 (83) is fulfilled iff $\sigma \leq \overline{\sigma}(\eta)$. It will be useful to define:

$$T(\eta) \equiv \frac{m_S^b + (1+i_S)(b_S^b + (\lambda + (1-\lambda)\eta)\ell_S^b) - (\lambda + \theta(1-\lambda)\eta)d_S^I}{1-\lambda}. \tag{84}$$

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 $T'(\eta) = (1+i_S)\ell_S^b - \theta d_S^I < 0$.[50] Substituting for $\overline{d}_R^P(\eta, \sigma)$, we can rewrite condition (83) as:

$$\begin{aligned} u(\eta q_S) &\geq \frac{(\eta+\sigma)}{(1-\theta(\eta+\sigma))}\frac{\phi}{1+\iota}T(\eta) \\ \Leftrightarrow \sigma\left[\theta u(\eta q_S) + \frac{\phi}{1+\iota}T(\eta)\right] &\leq (1-\theta\eta)u(\eta q_S) - \frac{\phi}{1+\iota}\eta T(\eta) \\ \Leftrightarrow \sigma &\leq \frac{u(\eta q_S)}{\theta u(\eta q_S) + \frac{\phi}{1+\iota}T(\eta)} - \eta \equiv \overline{\sigma}(\eta) \end{aligned} \tag{85}$$

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

$$\overline{\sigma}'(\eta) = \frac{\frac{\phi}{1+\iota}\big(q_S T(\eta)u'(\eta q_S) - u(\eta q_S)T'(\eta)\big)}{\left(\theta u(\eta q_S) + \frac{\phi}{1+\iota}T(\eta)\right)^2} - 1, \tag{86}$$

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

$$f' = \overbrace{\frac{\phi}{1+\iota}q_S^2 T(\eta)}^{>0}\,\overbrace{u''(\eta q_S)}^{<0} < 0 \tag{87}$$

$$g' = \underbrace{2\left[\theta u(\eta q_S) + \frac{\phi}{1+\iota}T(\eta)\right]}_{>0}\underbrace{\left[\theta q_S u'(\eta q_S) + \frac{\phi}{1+\iota}T'(\eta)\right]}_{>0 \text{ for } \eta\in[0,1]} > 0. \tag{88}$$

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 (88), note that

$$\begin{aligned} \theta q_S u'(\eta q_S) + \frac{\phi}{1+\iota}T'(\eta) &> 0 \\ \Leftrightarrow \theta q_S u'(\eta q_S) &> \frac{\phi}{1+\iota}(\theta d_S^I - (1+i_S)\ell_S^b) \\ \Leftrightarrow \frac{u'(\eta q_S)}{\rho_S} &> 1 - \frac{\phi}{1+\iota}\frac{(1+i_S)\ell_S^b}{\theta \rho_S q_S} \end{aligned} \tag{89}$$

where we used $T'(\eta) = -(\theta d_S^I - (1+i_S)\ell_S^b)$, $d_S^I = p_S q_S$ and $\rho_S \equiv \frac{\phi}{1+\iota}p_S$. Since $\frac{u'(q_S)}{\rho_S} > 1$ in steady state (see (32) and (34)), we know that condition (89) is fulfilled for any $\eta \in [0,1]$.

[50]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 $\theta d_S^I > (1+i_S)\ell_S^b$.

B.4.3 Threshold to Prevent a Run

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

$$\begin{aligned}(1-\lambda)(1-\theta(\eta+\sigma))d_S^I &\leq m_S^b+(1+i_S)(b_S^b+(\lambda+(1-\lambda)\eta)\ell_S^b)-(\lambda+\theta(1-\lambda)\eta)d_S^I)\\ \Leftrightarrow \sigma\theta(1-\lambda)d_S^I &\geq (\lambda+(1-\lambda))d_S^I-\left[m_S^b+(1+i_S)b_S^b+(1+i_S)(\lambda+(1-\lambda)\eta)\ell_S^b\right]\\ \Leftrightarrow \sigma &\geq \left[\frac{d_S^I-(m_S^b+(1+i_S)(b_S^b+\lambda\ell_S^b))}{(1-\lambda)\theta d_S^I}\right]-\frac{(1+i_S)\ell_S^b}{\theta d_S^I}\eta \equiv \underline{\hat{\sigma}}(\eta)\end{aligned} \tag{90}$$

Note that we have $\underline{\hat{\sigma}}(0)\in(0,1)$, and $\underline{\hat{\sigma}}'(\eta)=-\dfrac{(1+i_S)\ell_S^b}{\theta d_S^I}<0$, with $\left|\underline{\hat{\sigma}}'(\eta)\right|<1$.[51]

B.5 Proof of Proposition 10

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 B.4.1 and B.4.2 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 $\lim_{\eta\to0}\overline{\sigma}'(\eta)=+\infty$. 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

$$\begin{aligned}&\overline{\sigma}(\overline{\eta})>\underline{\sigma}(\overline{\eta})\\ \Leftrightarrow\quad&\frac{1}{\theta+\dfrac{\phi}{1+\iota}\dfrac{T(\overline{\eta})}{u(\overline{\eta}q_S)}}>\frac{\overline{\eta}(1-\lambda)d_S^I}{(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)+m_S^b-\lambda d_S^I}\\ \Leftrightarrow\quad&T(\overline{\eta})>\frac{\phi}{1+\iota}\frac{\overline{\eta}d_S^IT(\overline{\eta})}{u(\overline{\eta}q_S)}\\ &\frac{u(\overline{\eta}q_S)}{\overline{\eta}q_S}>\rho_S\end{aligned} \tag{91}$$

where we used $d_S^I=p_Sq_S\equiv\dfrac{1+\iota}{\phi}\rho_Sq_S$ 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 (32) and (34)), condition (91) is fulfilled, from which it follows that $\hat{\eta}^{max}>\overline{\eta}$.

Finally, from Appendix B.4.1, we know that the highest value of η consistent with $\underline{\sigma}(\eta)\leq1-\eta$, 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 10.

B.6 Example with $\lambda=0$ where Runs cannot be Prevented

Figure 3 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

[51]See Footnote 50 for why the absolute value of the derivative is less than one.

the example shown in Figure 2, and the parameter values are given by Table 2. As in the example of Figure 2, parameters imply that the economy is in a zero-lower bound equilibrium. We can see

Table 2: Parameter values for Figure 3.

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

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

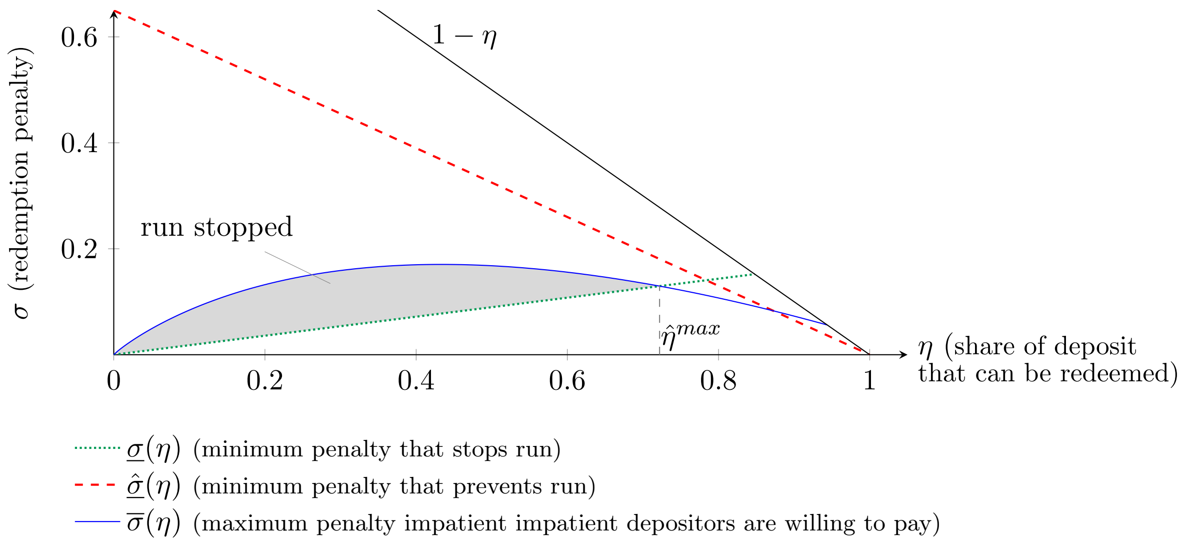

Figure 3: Example where runs cannot be prevented.

B.7 Proof of Proposition 11

From (63), we know that $\pi > \pi_S$ can only occur if $\chi < \delta_\ell$. From (40) and (59), we get that

$$\chi < \delta_\ell \;\Leftrightarrow\; \frac{\alpha(\rho_S)}{\alpha(\rho)} > \frac{\theta + \eta(1-\theta)}{\delta_\ell}. \tag{92}$$

Substituting for η using (58), 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 (92) 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

$$\pi > \pi_S \;\Rightarrow\; \rho\alpha(\rho)q^e(\rho) < \rho_S\alpha(\rho_S)q^e(\rho_S), \tag{93}$$

i.e., an increase in inflation implies a decrease in real loan repayments by active entrepreneurs. From (41), it follows that $\rho\alpha(\rho)q^e(\rho) = \max_{q^e}\{\rho q^e - c(q^e, k_S)\}$. 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 (93), this implies

$$\pi > \pi_S \;\Rightarrow\; \rho < \rho_S, \tag{94}$$

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

$$\frac{\alpha(\rho_S)}{\alpha(\rho)} = \frac{\rho q^e(\rho)}{\rho_S q^e(\rho_S)}\frac{1+\pi}{1+\pi_S}. \tag{95}$$

From (95) and the fact that $q^e(\rho)$ is strictly increasing in ρ (see the proof of Proposition 5), we get that $\pi > \pi_S$ implies

$$\frac{\alpha(\rho_S)}{\alpha(\rho)} < \frac{1+\pi}{1+\pi_S} = 1 + \frac{(1+i_S)\left(\delta_\ell - \frac{\alpha(\rho)}{\alpha(\rho_S)}(\theta + \eta(1-\theta))\right)\tau_\ell \ell_S^b - (1+i_S)(1-\delta_b)\tau_b b_S^b}{M_S + (1+i_S)B_S}, \tag{96}$$

i.e., the decline in the capital share relative to the steady state cannot exceed the increase in inflation. Note that we have used (40), (59) and (63) to substitute for $(1+\pi)/(1+\pi_S)$ in (96). The last step of the proof is to show that (92) and (96) are inconsistent with each other. To do this, we first take the derivative of the RHS in (96) with respect to $\alpha(\rho_S)/\alpha(\rho)$, which gives

$$\frac{\partial[(1+\pi)/(1+\pi_S)]}{\partial[\alpha(\rho_S)/\alpha(\rho)]} = \frac{(1+i_S)\tau_\ell \ell_S^b}{M_S + (1+i_S)B_S}\left(\frac{\alpha(\rho)}{\alpha(\rho_S)}\right)^2(\theta + \eta(1-\theta)) < 1, \tag{97}$$

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 (97) 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 (96) is violated for $\alpha(\rho_S)/\alpha(\rho) = [\theta + \eta(1-\theta)]/\delta_\ell$, i.e., if (92) holds at equality. Together with (97), this means there exists no $\alpha(\rho_S)/\alpha(\rho)$ that satisfies both (92) and (97). 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 (92) with a weak inequality, we still get a contradiction, such that $\delta_\ell < 1$ implies $\chi > \delta_\ell$.

B.8 Proof of Proposition 12

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 (57) and (58) 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

$$\frac{\rho\alpha(\rho)q^e(\rho)}{\rho_S\alpha(\rho_S)q^e(\rho_S)} \geq 1, \text{ with equality if } \chi < 1. \tag{98}$$

Using these results and (59), the fraction of active entrepreneurs satisfies:

$$\chi = \min\left\{\frac{[\theta+\eta(1-\theta)]\alpha(\rho)}{\alpha(\rho_S)}\frac{1+\pi_S}{1+\pi}, 1\right\} \tag{99}$$

Equation (60) remains unchanged, whereas Equation (61) (nominal liabilities of the government) becomes

$$d_S^I\{\theta+\eta(1-\theta)[\theta+\mathcal{I}(1-\theta)]\}+(1-\tau_b)(1+i_S)b_S^b-\tau_\ell\chi(1+i_S)\ell_S^b\frac{1+\pi}{1+\pi_S}+(1+i_S)(b_S-b_S^b). \tag{100}$$

Combining (60) with (100) and using $m_S^b = M_S$, we can express outstanding government liabilities after entrepreneurs repaid loans as

$$M_S+(1+i_S)\tau_\ell\ell_S^b\left(\delta_\ell-\chi\frac{1+\pi}{1+\pi_S}\right)+(1+i_S)(1-\tau_b(1-\delta_b))b_S^b+(1+i_S)(B_S-b_S^b), \tag{101}$$

so that equation (63) changes to

$$\frac{1+\pi}{1+\pi_S} = \max\left\{1+\frac{(1+i_S)\left(\delta_\ell-\chi\frac{1+\pi}{1+\pi_S}\right)\tau_\ell\ell_S^b-(1+i_S)(1-\delta_b)\tau_b b_S^b}{M_S+(1+i_S)B_S}, 1\right\}. \tag{102}$$

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 (102) that $\pi > \pi_S$ only if $\chi < 1$. From (98), we get that $\chi < 1$ only if $\rho = \rho_S$. From (99), we thus get that $\chi < 1$ implies $\chi = [\theta+\eta(1-\theta)]\frac{1+\pi_S}{1+\pi}$. Using this in condition (102), we get that $\pi > \pi_S$ only if $\delta_\ell > \theta+\eta(1-\theta)$. We have shown already in the proof of Proposition 11 that $\delta_\ell \leq \theta+\eta(1-\theta)$, so that we obtain a contradiction.