Flight to Liquidity and Systemic Bank Runs

Roberto Robatto,* University of Wisconsin-Madison

June 15, 2017

This paper presents a general equilibrium monetary model of fundamentals-based bank runs to study monetary injections during financial crises. When the probability of runs is positive, depositors increase money demand and reduce deposits; at the economy-wide level, the velocity of money drops and deflation arises. Two quantitative examples show that the model accounts for a large fraction of (i) the drop in deposits during the Great Depression and (ii) the $400 billion run on money market mutual funds in September 2008. In some circumstances, monetary injections have no effects on prices but reduce money velocity and deposits. Counterfactual policy analyses show that, if the Federal Reserve had not intervened in September 2008, the run on money market mutual funds would have been $141 billion smaller.

Keyworkds: Monetary Injections; Flight to Liquidity; Bank Runs; Endogenous Money Velocity; Great Depression; Great Recession; Money Market Mutual Funds.

1 Introduction

The bankruptcy of Lehman Brothers in September 2008 was followed by a flight to safe and liquid assets and runs on several financial institutions. For instance, Duygan-Bump et al. (2013) and Schmidt, Timmermann, and Wermers (2016) document a $400 billion run on money market mutual funds. In response to these events, the Federal Reserve implemented massive monetary interventions. Flight to liquidity, runs, and monetary interventions characterized the Great Depression as well, although the response of the Federal Reserve was more muted at the time, and the US economy experienced a large deflation (Friedman and Schwartz, 1963).

Despite the interactions between bank runs, flight to liquidity, and monetary policy interventions, very few models analyze the interconnections among these phenomena. Most of the literature on banking crises assumes that banks operate in environments with only one real good, without fiat money. While this approach is useful for many purposes, in practice banks take and repay deposits using money, giving rise to non-negligible interactions with monetary policy choices.[1]

*E-mail: robatto@wisc.edu. I am grateful to Gadi Barlevy, Saki Bigio, Elena Carletti, Briana Chang, Eric Mengus, Vincenzo Quadrini, João Ramos, David Zeke, and many other participants in seminars and conferences for their comments and suggestions. Kuan Liu has provided excellent research assistance.

[1] A few other papers deal with this observation. I review this literature in the next section.

To fill this gap, I present a general equilibrium model of fundamentals-based bank runs with money. If the fundamentals of the economy are strong, runs do not arise in equilibrium, and the outcomes in the banking sector look very similar to the good equilibrium in Diamond and Dybvig (1983). If instead the fundamentals of the economy are weak, the equilibrium is characterized by runs on many banks (i.e., systemic runs). Runs are associated with a flight to liquidity (i.e., an increase in money demand and a drop in deposits), deflation, a drop in nominal asset prices, and a drop in money velocity.

My objective is to use this model to study the effects of monetary injections on prices and quantities, especially financial and monetary variables, during systemic banking crises. I do not focus on welfare or optimal policy, although I briefly discuss these issues. Thus, the spirit of the main exercise is similar to the analysis of money in general equilibrium models with incomplete markets (see e.g. Magill and Quinzii, 1992) and the analysis of monetary policy shocks in monetary models (see e.g., Alvarez, Atkeson, and Edmond, 2009; Christiano and Eichenbaum, 1992; Rocheteau, Weill, and Wong, 2015).[2]

To highlight the mechanics and transmission mechanisms of monetary injections, I make some stark assumptions to keep the model simple and tractable. In particular, output is exogenous and there are no aggregate shocks, prices are fully flexible, and, in the baseline model, depositors' preferences are locally linear. In this way, my results can easily be compared with classical monetary models such as Lucas and Stokey (1987).

The main result of the paper is related to the analysis of temporary monetary injections, that is, injections that are reverted when the crisis is over. Temporary monetary injections produce unintended consequences during a crisis:[3] an amplification of the flight to liquidity (i.e., deposits drop in comparison to the economy without policy intervention) and a reduction in money velocity. Even though prices are fully flexible, they move less than one-for-one with the injection because of the endogenous reduction in velocity. Moreover, the amplification of the flight to liquidity is maximal when prices move the least. It is worth emphasizing that the same temporary injection implemented in an economy with strong

[2]The ultimate objective of the monetary policy shocks literature is to provide a validation of existing theories by comparing the effects of shocks in the model with those identified by vector autoregressions (VARs) in the data. Different, the objective of this paper is to provide a simple framework that explains the transmission mechanism of monetary policy during banking crises, which is an intermediate step in the ultimate goal of studying optimal policy and welfare. The literature is mostly silent on both the transmission mechanism and optimal policy, and I can only focus on the former due to space limitations.

[3]By "unintended consequences" I refer to the effects on prices and quantities that are not directly targeted by a policy intervention. As pointed out by Bernanke (2002) in his remarks on how to avoid deflation during financial crises, "each method of adding money [...] has advantages and drawbacks" and "calibrating the economic effects [...] may be difficult, given our [...] lack of experience with such policies."

fundamentals and no runs produces standard effects and no unintended consequences. I argue that these findings are important for the analysis of actual financial crises because several monetary policy interventions implemented during both the Great Depression and the Great Recession are best characterized as temporary. I first derive the results theoretically, and then I present two quantitative examples applied to the Great Depression and the Great Recession, showing that the channel identified by my model is economically important. Therefore, my analysis goes one step further than most bank runs papers that use microfounded models, which typically focus only on qualitative studies.[4]

The unintended consequences of temporary monetary injections are related to the role of money in the model. To understand this role, recall first the structure of typical three-period bank runs models ($t = 0, 1, 2$) without money, such as Diamond and Dybvig (1983), Allen and Gale (1998), and Goldstein and Pauzner (2005). In these models, households deposit all their wealth into banks at $t = 0$. This is the case no matter whether depositors assign, at $t = 0$, zero probability to runs at $t = 1$ (as in Diamond and Dybvig, 1983) or a positive probability (as in Allen and Gale, 1998, and Goldstein and Pauzner, 2005). In contrast, there is an explicit role for fiat money in my model, and households deposit all their money at $t = 0$ only if the probability of a run is zero. If the probability of runs is instead positive, households keep some money in their wallets. In this case, households' money demand depends on its opportunity cost, which is represented by the nominal return paid by productive assets.

To understand the transmission channel of a temporary monetary injection, it is useful to deconstruct this policy into two separate interventions. A temporary monetary injection is the "sum" of (i) a permanent monetary injection implemented during a crisis and (ii) a permanent reduction of money supply, of the same size, implemented when the crisis is over. Crucially, the second intervention is fully *anticipated* because the central bank announces a temporary injection to begin with. This anticipation has an impact on the flight to liquidity by affecting the opportunity cost of holding money.

A permanent monetary injection implemented during a crisis has standard effects. That is, this intervention does not affect velocity, and thus, current and future prices increase one-for-one with the injection (both nominal asset prices and the price level).

A permanent reduction of money after the crisis reduces prices after the crisis. If this intervention were completely unanticipated, there would be no additional effects. However,

[4] A few recent papers use microfounded models of bank runs for quantitative analyses, such as Angeloni and Faia (2013), Egan, Hortaçsu, and Matvos (2017) and Gertler and Kiyotaki (2015).

the permanent reduction of money after the crisis is anticipated, and thus, it also produces effects before its implementation, while the crisis is still unfolding: a reduction of prices, a reduction of money velocity, and an amplification of the flight to liquidity. To understand these results, note that a future reduction of money reduces future prices – in particular, future nominal asset prices. This effect creates a downward pressure on the nominal return on productive assets, that also represents the opportunity cost of holding money. As a result of the lower opportunity cost, households hold more money during the crisis by reducing deposits; that is, they amplify the flight to liquidity. In addition, velocity and prices drop during the crisis because of the negative relationship between the flight to liquidity on the one hand and velocity and prices on the other.[5]

Importantly, temporary injections in non-crisis times do not produce any unintended consequences and have standard effects; that is, prices move one-for-one with monetary injections and deposit decisions are unchanged. Without runs, households deposit all their money at banks, and thus, hold no money in their wallets to economize on the opportunity cost of holding money. In this context, a change in the opportunity cost of money triggered by policy interventions does not alter households' decision to hold no money in their wallets, provided that such cost remains positive. The distinction between crisis and non-crisis times is reminiscent of Magill and Quinzii (1992), in which the real effects of monetary policy depend on whether agents store money or not.

A key element that governs the magnitude of the response to a temporary monetary injection is households' elasticity of money demand with respect to its opportunity cost. The baseline model features a very high money demand elasticity due to some stark assumptions. In response to a temporary monetary injection, the high elasticity produces a substantial amplification of the flight to liquidity; moreover, all the effects on prices described above offset each other, and thus asset prices and the price level are constant. That is, the baseline model is characterized by a very high degree of monetary non-neutrality. I then analyze the robustness of the results to a model with standard preferences, though I have to rely on numerical analysis. Under these preferences, the elasticity of money demand and the degree of monetary non-neutrality are lower, but the channel that drives the results is unchanged. Temporary monetary injections still reduce velocity and, depending on the size of the monetary injection, amplify the flight to liquidity. In addition, temporary

[5]The negative relationship between the flight to liquidity on the one hand and velocity and prices on the other is exemplified by the fact that velocity and prices are low when fundamentals are weak and depositors fly to liquidity, whereas velocity and prices are high when fundamentals are strong and depositors do not fly to liquidity.

monetary injections increase nominal prices (both asset prices and the price level) but do so less than one-for-one because of the endogenous reduction in velocity.

I emphasize that, in the baseline model, temporary injections change depositors' behavior even though they do not affect equilibrium prices. This consideration provides a second, more formal way of understanding the main results. In the baseline model, households have locally linear preferences and thus they are indifferent among several choices as long as prices are such that their first-order conditions hold with equality. Given such prices, the market clearing conditions can then be used to solve for equilibrium quantities. Therefore, the transmission of temporary injections works through the market clearing conditions rather than prices. Focusing on the limit case with locally linear preferences is useful because it rules out incorrect intuitions. Even in the model with standard preferences, in which temporary injections do have an impact on prices, the amplification of the flight to liquidity is maximal in scenarios in which equilibrium prices are affected the least.

Finally, I consider two quantitative examples based on the model with standard preferences: one for the 2008 crisis and one for the Great Depression. Let me emphasize that the model is deliberately simple and abstracts from other forces that might be at work in richer frameworks. Nonetheless, abstracting from such forces allows me to isolate the magnitude of the channel that I have identified.

The model accounts for about 40% of the drop in deposits during the Great Depression and for a similar fraction of the $400 billion redemptions from money market mutual funds during the run that took place in September 2008. The policy analyses show that if the Federal Reserve had temporarily injected an extra dollar during the Great Depression or the Great Recession, it would have substantially amplified the flight to liquidity, with little effects on nominal prices. Moreover, I ask what would have happened in 2008 if the Federal Reserve had not set up facilities to provide liquidity to mutual funds. The model predicts that deflation would have occurred but the run would have been $141 billion smaller. According to the model, the Federal Reserve avoided deflation in 2008 at the expense of an amplification of runs and of the flight to liquidity.

1.1 Additional comparisons with the literature

A few other papers analyze monetary injections in the context of bank runs. However, these papers differ from mine in important ways.

A first set of papers analyze monetary injections in the context of bank runs driven by fundamentals. Allen and Gale (1998), Allen, Carletti, and Gale (2013), and Diamond and Rajan (2006) study how monetary policy should respond to aggregate shocks when deposit contracts are nominal and not contingent on the price level. However, crises in these models do not produce flight to liquidity in anticipation of runs or deflation, and small monetary injections may actually generate inflation. As a result, the main focus of these papers is on other aspects of banking crises.[6] Antinolfi, Huybens, and Keister (2001) and Rochet and Vives (2004) study central bank lending in response to aggregate shocks in models that do not produce any flight to liquidity either.

A second set of papers present models in which monetary injections can eliminate bank runs driven by panics, in the sense of multiple equilibria. Carapella (2012), Cooper and Corbae (2002), and Robatto (2015) use general equilibrium models, whereas Martin (2006) analyzes a Diamond-Dybvig partial-equilibrium economy with money. I comment further on the two closest papers, Cooper and Corbae (2002), and Robatto (2015). In Cooper and Corbae (2002), depositors choose to hold some money in their wallets during crises, as in my model. However, they focus solely on steady states in which banks are either perpetually well functioning or malfunctioning, and thus they consider only permanent injections. In contrast, my simpler model allows me to study a scenario in which crises eventually end and to distinguish between temporary and permanent injections. In Robatto (2015), I build an infinite-horizon, monetary model of bank runs driven by panics. In some circumstances, temporary monetary injections produce some unintended consequences as well. However, the richness of that model – required to study multiple equilibria in an infinite-horizon economy – imposes limitations on the analysis. Moreover, the focus of Robatto (2015) is on the monetary policy stance that eliminates multiple equilibria, similar to the main research question in Carapella (2012) and Cooper and Corbae (2002).

2 Baseline model: the core environment

This section presents the core environment without banks, and Section 3 derives the equilibrium. Sections 4 and 5 extend this core environment by introducing banks. The objective

[6] Allen and Gale (1998) and Allen, Carletti, and Gale (2013) emphasize that nominal deposit contracts allow the economy to achieve the first best in response to aggregate shocks under an appropriate monetary intervention. In contrast, in my model, there are no aggregate shocks and the denomination of deposits does not play any role. Diamond and Rajan (2006) emphasize the comparison between deposits denominated in foreign versus domestic currency. They also sketch an extension of their model in which runs are associated with deflation; however, they do not analyze monetary injections in the extended model with deflation.

is to present a simple framework that allows me to explain the intuition of the unintended consequences of monetary injections. Section 6 presents a richer framework that relaxes some of the assumptions used in the baseline model, showing that the main forces are still at work and can be quantitatively relevant.

Time is discrete with three periods indexed by $t \in \{0, 1, 2\}$. The economy is populated by a double continuum of households indexed by $h \in \mathbb{H} = [0, 1] \times [0, 1]$; the double continuum is required when introducing banks in Section 4.

The core environment combines preference shocks at $t = 1$, in the spirit of Diamond and Dybvig (1983), with a Lucas-tree cash-in-advance economy. Cash is required to finance consumption expenditure at $t = 1$, after agents are hit by preference shocks. As a result, a precautionary demand for money arises at $t = 0$, so that households can finance consumption induced by preference shocks at $t = 1$. In order to deal with money in a finite-horizon model, I introduce a technology to transform money into consumption goods at $t = 2$; in a related paper (Robatto, 2015), I present an infinite-horizon model of banking that motivates this assumption. That is, money has a continuation value because it can be carried over to the next period.

2.1 Preferences

Let C_1^h and C_2^h denote consumption of household h at $t = 1$ and $t = 2$, respectively. Households' utility depends on a preference shock that is realized at the beginning of $t = 1$:

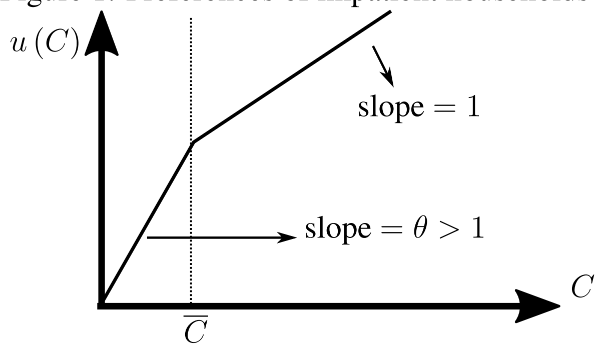
$$\text{utility} = \begin{cases} u\left(C_1^h\right) + \beta C_2^h & \text{(impatient household)} \quad \text{with probability } \kappa\,. \\ \beta\, C_2^h & \text{(patient household)} \quad \text{with probability } 1 - \kappa \end{cases} \tag{1}$$

Note that both patient and impatient households derive linear utility from consumption at $t = 2$. The function $u(\cdot)$ is piecewise-linear, as represented in Figure 1:

$$u\left(C_1^h\right) = \begin{cases} \theta C_1^h & \text{if } C_1^h < \overline{C} \\ \theta \overline{C} + \left(C_1^h - \overline{C}\right) & \text{if } C_1^h \geq \overline{C} \end{cases} \qquad \theta > 1,\ \overline{C} > 0. \tag{2}$$

The assumption $\theta > 1$ captures impatience. If consumption at $t = 1$ is $C_1^h < \overline{C}$, the marginal utility at $t = 1$ is $\theta > 1$ and thus larger than the marginal utility at $t = 2$, which equals one. If instead $C_1^h \geq \overline{C}$, both marginal utilities are one. This structure gives rise to an important driving force, namely, a desire to consume at least $\overline{C}$ if h is impatient.

Figure 1: Preferences of impatient households



Another way to understand the role of $\theta > 1$ is to note that $u(\cdot)$ is globally concave, and thus households are (globally) risk averse with respect to time-1 consumption.

The local linearity delivers neat closed-form outcomes. Nonetheless, the main results are robust to a more standard smooth utility function. In this case, though, some analyses can be performed only numerically. More discussion is provided in Section 6.

The preference shock is i.i.d. across households, and I assume that the law of large numbers holds, so that the fraction of impatient agents in the economy equals κ. Moreover, I assume that the law of large numbers also holds for each subset of $\mathbb{H}$ with a continuum of households.[7] The preference shock is private information of household h.

2.2 Assets, production and markets

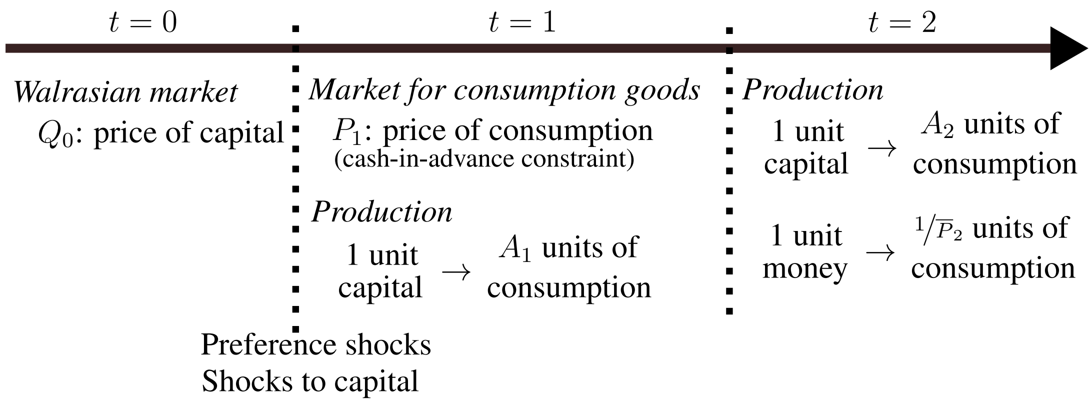
There are two assets with exogenous supply: money and capital. The initial endowment of money at the beginning of $t = 0$ (which can be understood as depending on monetary policy choices in an unmodeled date -1) is given by $\overline{M}$. The money supply at $t = 0, 1, 2$ is denoted by M_t^S and is controlled by the central bank. In this economy without banks, I consider a constant money supply, $M_t^S = \overline{M}$ for $t = 0, 1, 2$, whereas in Section 4 I describe how the central bank can vary the money supply. Money is the numeraire. Without loss of generality, contracts are expressed in terms of money as well.[8] Capital is in fixed supply $\overline{K}$. The fixed-supply assumption is made for convenience because it permits abstracting from endogenous investment decisions.[9]

[7] This is consistent with the results of Al-Najjar (2004) about the law of large numbers in large economies.

[8] Different from Allen, Carletti, and Gale (2013), the denomination of contracts is irrelevant in my model because there are no aggregate shocks; that is, the results are unchanged if deposits were contingent on prices.

[9] The results would remain unchanged if I were to endow households with goods rather than a fixed supply of capital. In this case, the entire endowment would be invested anyway because there is no consumption at $t = 0$. That is, I follow the approach of monetary models similar to Lucas and Stokey (1987), in which

Figure 2: Timing of production and markets



Capital is hit by idiosyncratic, uninsurable shocks at $t = 1$. The effect of these shocks is to reallocate capital among agents, leaving the aggregate stock of capital unchanged at $\overline{K}$.[10] For a fraction $\alpha \in (0, 1)$ of agents, the stock of capital reduces by a factor of $1 + \psi^L$, where $-1 \leq \psi^L \leq 0$; that is, if a household bought K_0^h capital at $t = 0$ and is hit by ψ^L, its stock of capital at $t = 1$ is $K_0^h \left(1 + \psi^L\right)$. For the other $1 - \alpha$ agents, capital increases by a factor of $1 + \psi^H$, where $\psi^H \geq 0$. Since the shocks are idiosyncratic, they satisfy

$$\alpha \left(1 + \psi^L\right) + (1 - \alpha) \left(1 + \psi^H\right) = 1 \tag{3}$$

Without loss of generality, I set $\psi^L = -1$, so that an agent hit by ψ^L loses all its capital stock. The results are unchanged if $0 < \psi^L \leq -1$, due to the risk neutrality of households with respect to time-2 consumption. However, setting $\psi^L = -1$ simplifies the exposition and the analysis. In the rest of the paper, I use α to describe the stochastic process of the idiosyncratic shocks, whereas ψ^H is determined residually by Equation (3). The idiosyncratic shocks do not play a major role in the bankless economy but are crucial to produce runs in the economy with banks.

Next, I describe trading and production. The timing is represented in Figure 2.

At $t = 0$, there is a Walrasian market in which capital and money can be traded. The price of capital is denoted by Q_0.

At $t = 1$ (after preference shocks and capital shocks are realized), each unit of capital produces A_1 units of consumption goods that can be sold at price P_1. Consumption

portfolio decisions involve holding money or physical capital, whereas perishable goods cannot be stored.

[10] Alternatively, these disturbances could be modeled as idiosyncratic shocks that affect the productivity of capital at $t = 1$ and $t = 2$.

expenditures are subject to a cash-in-advance constraint; as in Lucas and Stokey (1987), households cannot consume goods produced by their own stock of capital. Capital is illiquid at $t = 1$, that is, it cannot be traded. This restriction is required to provide a role for banks, similar to Jacklin (1987); if households could trade capital at $t = 1$ and use the proceeds of trade to consume, there would be no role for banks.

At $t = 2$, each unit of capital produces A_2 units of consumption goods and each unit of money produces ${}^{1}\!/\overline{P}_2$ units of consumption goods; the parameter $\overline{P}_2$ is exogenous but is motivated by an infinite-horizon formulation in which fiat money can be carried over and used in the next period. Thus, total consumption available at $t = 2$ is $A_2\overline{K} + \left(1/\overline{P}_2\right) M_2^S$. In Section 2.4, I impose a restriction on $\overline{P}_2$ to make sure that the central bank cannot increase consumption by printing money.

For future reference, let $1 + r_2^K(\psi)$ be the nominal return on capital at $t = 2$ for an agent that is hit by the idiosyncratic shock to capital ψ. This return is defined by

$$1 + r_2^K(\psi) = (1 + \psi)\,\frac{A_2\overline{P}_2 + A_1 P_1}{Q_0}. \tag{4}$$

2.3 Endowments

Without loss of generality, I assume that all households have the same endowment of money and capital at $t = 0$. Thus, each household h is endowed with money $\overline{M}$ and capital $\overline{K}$.

2.4 Restrictions on parameters

I impose a restriction on the parameters A_2 and $\overline{P}_2$ that govern the output produced by capital and money at $t = 2$:

$$A_2 = \frac{\beta}{1-\beta} A_1, \qquad \overline{P}_2 = \frac{M_2^S}{A_1\overline{K}}\,. \tag{5}$$

These restrictions imply that period $t = 2$ in the three-period model can be interpreted as period "$t + 1$" in an infinite-horizon economy. First, the restriction on A_2 would imply, in an infinite-horizon economy, that the real value of capital in "$t + 1$" is equal to the present-discounted value of future dividends. Second, the value of $\overline{P}_2$ would imply money neutrality in "$t + 1$"; that is, in an infinite-horizon model, the price level $\overline{P}_2$ would increase one-to-one with the money supply M_2^S.[11]

In my three-period model, the restriction on $\overline{P}_2$ implies that the central bank cannot increase consumption at $t = 2$ by printing money. To see this, recall that output available for consumption at $t = 2$ is $A_2\overline{K} + \left(1/\overline{P}_2\right) M_2^S$, which equals $A_1\overline{K}/\left(1-\beta\right)$ using (5). That is, total consumption at $t = 2$ is independent of the money supply at $t = 2$.

I also impose a restriction on the parameter $\overline{C}$ that governs the utility of impatient households defined in (2):

$$\overline{C} = \frac{A_1\overline{K}}{\kappa}. \tag{6}$$

The $A_1\overline{K}/\kappa$ is the level of consumption at $t = 1$ that can be achieved if all impatient households consume the same amount (total production at $t = 1$ is $A_1\overline{K}$, and there is a mass κ of impatient agents). Equation (6) implies that there is a feasible allocation in which the consumption of impatient households is equalized at $\overline{C}$, and thus their marginal utility equals one; that is, no impatient household has marginal utility $\theta > 1$ in this allocation. For technical reasons, some results require the utility function $u\left(C\right)$ to be differentiable at $C = A_1\overline{K}/\kappa$ and its derivative to equal one. To guarantee these results, Equation (6) can be replaced with $\overline{C} = A_1\overline{K}/\kappa - \xi$, with $\xi > 0$ but arbitrarily small.

Finally, the discount factor β satisfies $\beta < 1$ and is sufficiently close to one.

3 Baseline bankless economy: results

I now study the equilibrium of the economy presented in Section 2. Since households are the only set of private agents in the economy and there are no banks, I refer to this environment as the bankless economy.

Households choose money M_0^h, capital K_0^h, and consumption C_1^h and C_2^h by solving

$$\max_{M_0^h, K_0^h, C_1^h} \kappa \left\{ u\left(C_1^h\right) + \beta \underbrace{\frac{\left(M_0^h - P_1 C_1^h\right) + Q_0 K_0^h \, \mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2}}_{=C_2^h \text{ if } h \text{ is impatient}} \right\}$$
$$+ \left(1-\kappa\right)\beta \underbrace{\frac{M_0^h + Q_0 K_0^h \, \mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2}}_{=C_2^h \text{ if } h \text{ is patient}} \tag{7}$$

[11] In particular, in an infinite-horizon economy, the value of $\overline{P}_2$ in (5) would arise in a steady state in which banks are active and there are no runs. In addition, note that the expression for $\overline{P}_2$ in (5) is similar to that derived for P_1 in Proposition 5.1.

where the expectation is taken with respect to the shocks to capital held by agent h, ψ^h. The maximization in (7) is subject to the budget and cash-in-advance constraints:

$$M_0^h + K_0^h Q_0 \leq \underbrace{\overline{M} + \overline{K} Q_0}_{\text{value of endowments}} \tag{8}$$

$$P_1 C_1^h \leq M_0^h\,. \tag{9}$$

In (7), I use the fact that the optimal consumption of patient households at $t = 1$ is zero, and thus C_1^h refers to the consumption at $t = 1$ if the household is impatient. At $t = 0$, the household has access to the Walrasian market where it can adjust its portfolio of money and capital, subject to the budget constraint (8); M_0^h and K_0^h denote the amount of money and capital that the household has after trading. At $t = 1$, consumption is subject to the cash-in-advance constraint (9). At $t = 2$, consumption is financed with unspent money ($M_0^h - P_1 C_1^h$ if the household is impatient and M_0^h if it is patient) and capital bought at $t = 0$ plus its return $r_2^K\left(\psi^h\right)$. The return on capital includes the proceeds from selling output $A_1 K_0^h$ (produced by capital at $t = 1$) at price P_1 and the output produced at $t = 2$.

To solve problem (7), I conjecture that the cash-in-advance constraint (9) holds with equality for impatient households. This conjecture is verified later because the opportunity cost of holding money, represented by the expected return on capital $\mathbb{E}\left\{r_2^K\left(\psi\right)\right\}$, is positive in equilibrium. Thus, it is not optimal for households to carry money that will be unspent.

The first-order conditions imply the following:

$$\beta\,\mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}\frac{1}{\overline{P}_2} = \kappa\,u'\left(C_1^h\right)\frac{1}{P_1} + (1-\kappa)\,\beta\frac{1}{\overline{P}_2}. \tag{10}$$

Households are indifferent between investing an extra dollar in capital or in money at $t = 0$. Investing in capital gives a return $\mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}$, discounted by the factor β and evaluated in units of time-2 consumption (i.e., the return is multiplied by $1/\overline{P}_2$). Investing in money allows households to increase consumption at $t = 1$ if the household is impatient (i.e., with probability κ) or at $t = 2$ if the household is patient (i.e., with probability $1 - \kappa$).

An equilibrium of this economy is a collection of prices Q_0 and P_1 and households' choices M_0^h, K_0^h, C_1^h, such that (i) M_0^h, K_0^h, and C_1^h solve the problem (7) given prices, (ii) the money and capital markets clear at $t = 0$, $\overline{M} = \int M_0^h dh$ and $\overline{K} = \int K_0^h dh$, and (iii) the goods market clears at $t = 1$, $\int C_1^h dh = A_1 \overline{K}$.[12]

[12]More generally, the market-clearing condition for money can be stated as $M_0^S = \int M_0^h dh$; however, the

In equilibrium, all households are alike at $t = 0$, and thus market clearing implies that they hold the same amount of money and capital, $M_0^h = \overline{M}$ and $K_0^h = \overline{K}$ for all h. At $t = 1$, only impatient households consume; since there is a mass κ of them and total output is $A_1\overline{K}$, consumption is $C_1^h = A_1\overline{K}/\kappa$.

Next, I solve for the price level at $t = 1$. The money spent is $\kappa\overline{M}$ because only impatient households spend money and consume at $t = 1$ (the mass of impatient households is κ, and each of them holds $M_0^h = \overline{M}$ money). The consumption expenditure is $P_1 \int C_1^h dh = P_1 A_1 \overline{K}$, where the equality follows from the goods market clearing at $t = 1$. Equating the money spent with the consumption expenditure, I can solve for the price level P_1:

$$P_1 = \frac{\kappa\overline{M}}{A_1\overline{K}}. \tag{11}$$

To solve for Q_0, I first use Equation (6) to note that the consumption of impatient households is at the kink of the utility function, $C_1^h = \overline{C}$, and thus the marginal utility of any additional unit of consumption is one: $u'\left(C_1^h\right) = 1$. Plugging $u'\left(C_1^h\right) = 1$ and Equation (11) into Equation (10), I can solve for the expected return on capital $\mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}$ and, using Equation (4), for the price of capital Q_0. The results are $\mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\} = \frac{1}{\beta}\left[1 + (1-\kappa)\beta\right]$ and $Q_0 = \frac{\beta}{1-\beta}\frac{\overline{M}}{\overline{K}}\left[\frac{\kappa+\beta(1-\kappa)}{1+\beta(1-\kappa)}\right]$. Consumption at $t = 2$ is $C_1^h = A_1\overline{K}\left[\frac{\beta}{1-\beta} + \kappa\right]$ if h is impatient and $C_1^h = A_1\overline{K}\left[\frac{\beta}{1-\beta} + 1 + \kappa\right]$ if h is patient.

Finally, I comment on welfare. At $t = 1$, the consumption of impatient households is equalized at $C_1^h = A_1\overline{K}/\kappa$, whereas the consumption of patient households is zero. This allocation is the same as what a social planner would choose. That is, banks have no role in increasing welfare in this baseline model. Nonetheless, two important remarks are in order. First, introducing banks results in significant effects on equilibrium prices and policy analysis. The existence of banks, which provide deposits that allow households to withdraw at $t = 1$ or to receive a return at $t = 2$, affects money demand and prices and has crucial implications for the transmission of monetary policy. Second, the main argument of the paper about to the effects of monetary injections on prices and deposits is independent of the welfare analysis. The simplest way to convey the main argument, then, is to use this baseline model. In the richer model of Section 6, welfare in the bankless economy is less than in the first best, and there is a welfare-increasing role for banks. Nonetheless, the main message about the effects of temporary monetary injections is unchanged.

supply of money in this economy is constant, and thus $M_0^S = \overline{M}$.

4 Baseline model with banks

I now extend the core environment of Section 2 by introducing a unit mass of banks indexed by b that act competitively and a central bank that can change the money supply. Similar to the previous sections, I use the superscript b to denote variables that refer to bank b.

Depending on parameters that govern the fundamentals of the economy, the equilibrium has either no runs at $t = 1$ or runs on some banks at $t = 1$. Therefore, runs are driven by fundamentals, as in Allen and Gale (1998), rather than by panics, as in Diamond and Dybvig (1983).[13]

The interaction between banks and households is standard. Households' endowments are the same as in Section 2, whereas banks have no endowment. Each bank is associated with a unit continuum of households and takes prices as given.[14] Households deposit at $t = 0$ and have the possibility to withdraw money at $t = 1$. If a household does not withdraw at $t = 1$, its deposits are repaid at $t = 2$ with a return, which can be positive (if the bank is solvent) or negative (if the bank is insolvent).

Recall that capital is subject to shocks at $t = 1$, and thus capital held by banks is hit by these shocks as well. I denote ψ^b to be the shock to capital held by bank b; I continue to denote ψ^h to be the shock to capital held by household h. As in the bankless economy, ψ^h does not play a major role because households are risk neutral at $t = 2$. In contrast, ψ^b is crucial because a bank becomes insolvent and is subject to a run if it is hit by the negative shock ψ^L.

4.1 Budgets and interaction between households and banks

$t = 0$: trading and deposits. Bank b buys money M_0^b and capital K_0^b using deposits D_0^b:

$$\underbrace{K_0^b Q_0}_{\text{capital}} + \underbrace{M_0^b}_{\text{money}} \leq \underbrace{D_0^b}_{\text{deposits}} \tag{12}$$

subject to the non-negativity constraints $M_0^b \geq 0$, $K_0^b \geq 0$, and $D_0^b \geq 0$.

[13] The lack of market for capital at $t = 1$ prevents banks from liquidating their assets at that time. This modeling assumption shuts the channel that gives rise to multiple equilibria in Diamond and Dybvig (1983), as noted by Jacklin and Bhattacharya (1988).

[14] Since there is a unit mass of banks and each bank is associated with a unit continuum of households, there is a well-defined link between the unit mass of banks and the double continuum of households introduced in Section 3.

Banks' allocation of deposits D_0^b between money M_0^b and capital K_0^b is the only relevant choice taken by banks. The other modeling assumptions related to banks and introduced later imply that repayment to households by banks at $t = 1$ and $t = 2$ depends only on the allocation of deposits across money and capital at $t = 0$.

Household h makes its portfolio decisions by choosing money, deposits, and capital:

$$\underbrace{M_0^h}_{\text{money}} + \underbrace{D_0^h}_{\text{deposits}} + \underbrace{Q_0 K_0^h}_{\text{capital}} \leq \underbrace{\overline{M} + \overline{K} Q_0}_{\text{value of endowments}} \tag{13}$$

subject to the non-negativity constraints $M_0^h \geq 0$, $D_0^h \geq 0$, and $K_0^h \geq 0$.

Each household can hold its deposits D_0^h only at one bank. This assumption can be justified by the costs of maintaining banking relationships. Formally, the cost is zero if household h holds deposits at one bank and infinite if household h holds deposits at two or more banks. This assumption implies that households face the risk that their own bank may be hit by the negative shock ψ^L and subject to a run. If households could deposit at all banks, they would diversify away this risk. The results are unchanged if households can deposit at, for example, two or three banks, but it is crucial that households cannot hold deposits at a large number of banks.

$t = 1$: withdrawals and consumption. Households observe their preference shocks and then decide their withdrawals, W_1^h. For each bank b, total withdrawals by its depositors cannot exceed the amount of money M_0^b chosen at $t = 0$ by the bank:

$$W_1^b = \int_{\left\{\substack{\text{depositors} \\ \text{of bank } b}\right\}} W_1^h dh \leq M_0^b, \tag{14}$$

where the integral is taken with respect to households that hold deposits at bank b. The inequality in (14) must hold because capital cannot be liquidated at $t = 1$.

Three assumptions govern withdrawals at $t = 1$. Appendix A provides further discussion about the role that each assumption plays in affecting the equilibrium.

a. At each bank, withdrawals are repaid based on a sequential service constraint;
b. Each bank has to repay the full value of deposits that are demanded back at $t = 1$ as long as the bank has money available. That is, if an household demand withdrawals $W_1^h = D_0^h$ at $t = 1$ and the bank has money in its vault when the household is served, the bank has to repay $W_1^h = D_0^h$. In other words, no haircut on deposits or no

suspension of convertibility can be imposed at $t = 1$;

c. Households' withdrawals at $t = 1$ can take values $W_1^h \in \left\{0, D_0^h\right\}$. That is, a household can either withdraw all its deposits at $t = 1$, $W_1^h = D_0^h$ (if it is served when a bank has money in its vault), or withdraw no money and wait until $t = 2$, $W_1^h = 0$, but it cannot choose to withdraw any amount in between.

Items (a) and (b) give rise to a limit on withdrawal determined by the position in line during a run. In the event of a run, households at the beginning of the line can withdraw all their deposits, but those at the end of the line cannot withdraw any money, because the bank does not have enough cash to serve them. Combining Items (a) and (b) with the assumption in Item (c), withdrawals are then given by the following:

$$W_1^h = \begin{cases} D_0^h & \text{if there is no run, or if } h \text{ is at the beginning of the line in a run} \\ 0 & \text{if } h \text{ is at the end of the line in a run.} \end{cases}$$

The fraction of households that are able to withdraw $W_1^h = D_0^h$ depends on the bank's investment in money at $t = 0$, M_0^b. Given deposits D_0^b, the higher are money holdings M_0^b, the higher is the fraction of depositors that are able to withdraw in the event of a run.

After making withdrawals, households choose consumption expenditure $P_1 C_1^h$ subject to a cash-in-advance constraint; that is, $P_1 C_1^h$ cannot exceed the sum of money M_0^h (chosen at $t = 0$) and withdrawals W_1^h:

$$P_1 C_1^h \leq M_0^h + W_1^h. \tag{15}$$

***t* = 2: return on deposits and consumption.** At $t = 2$, banks are liquidated, and the proceeds are used to pay deposits that have not been withdrawn at $t = 1$. Let $1 + r_2^b\left(\psi^b\right)$ denote the return on deposits not withdrawn. This return is possibly bank-specific because it depends on banks' choices of money and deposits made at $t = 0$, M_0^b and D_0^b, and by the idiosyncratic shock to capital ψ^b.

I focus on the relevant case in which all the money M_0^b is withdrawn by depositors at $t = 1$. In this case, the return $r_2^b\left(\psi^b\right)$ is paid using the return on capital bought at $t = 0$ and

$$r_2^b\left(\psi^b\right) = \begin{cases} r_2^K\left(\psi^H\right) & \text{if } \psi^b = \psi^H \\ -1 & \text{if } \psi^b = \psi^L. \end{cases} \tag{16}$$

The return $r_2^K\left(\psi^H\right)$ is defined in Equation (4), whereas the second entry of (16) follows from the assumption $\psi^L = -1$. That is, for a bank hit by ψ^L at $t = 1$ (and under the conjecture that all the money is withdrawn at $t = 1$), there are no resources left at $t = 2$, and thus, deposits not withdrawn at $t = 1$ are completely lost. The fact that deposits not withdrawn at $t = 1$ are completely lost triggers a run on such banks in equilibrium.

After deposits are repaid at $t = 2$, households consume C_2^h. Similar to the bankless economy, capital bought at $t = 0$, K_0^h, its return $r_2^K\left(\psi^h\right)$, and unspent money are used to finance consumption. In addition, consumption is also financed by deposits not withdrawn $D_0^h - W_1^h$ plus the return $r_2^b\left(\psi^b\right)$ paid by the bank, and lump-sum transfer T_2 from the central bank, if any (see Section 4.2). Therefore, household consumption at $t = 2$ is

$$C_2^h = \underbrace{\frac{Q_0 K_0^h \left[1 + r_2^K\left(\psi^h\right)\right]}{\overline{P}_2}}_{\text{capital + return}} + \frac{1}{\overline{P}_2}\left[\underbrace{\left(D_0^h - W_1^h\right)\left(1 + r_2^b\left(\psi^b\right)\right)}_{\text{deposits not withdrawn + return}} + \underbrace{\left(M_0^h + W_1^h - P_1 C_1^h\right)}_{\text{unspent money}} + T_2\right]. \tag{17}$$

4.2 Central bank

(Readers only interested in the model without policy intervention can skip this section.) Recall that money supply is denoted by M_t^S, $t = 0, 1, 2$, and is controlled by the central bank. If there is no policy intervention, the money supply is constant at $M_t^S = \overline{M}$ for all t.

If there is a policy intervention, the central bank changes the money supply by varying M_t^S. Interventions are announced at $t = 0$, before the Walrasian market opens; the central bank fully commits to the policy announcement.

If the money supply at $t = 0$ is $M_0^S > \overline{M}$, the central bank is injecting $M_0^S - \overline{M}$ units of money at $t = 0$ because the initial endowment of money is $\overline{M}$. The monetary injection is delivered using *asset purchases*, that is, purchases of capital K_0^{CB} on the market at price Q_0. The budget constraint of the central bank at $t = 0$ is

$$\underbrace{Q_0 K_0^{CB}}_{\text{asset purchases}} \leq M_0^S - \overline{M}. \tag{18}$$

The main results are unchanged if the central bank uses the newly printed money, $M_0^S - \overline{M}$, to offer loans to banks b, as long as such loans are fully collateralized using capital. Buying capital directly is equivalent to offering loans that are used by banks b to buy capital, which is in turn offered as collateral with the central bank.

At $t = 1$, I restrict attention to the case in which the money supply does not change because there is no market in which capital can be traded. Thus, $M_1^S = M_0^S$.

At $t = 2$, the central bank can again change the money supply. Monetary injections at $t = 2$ are implemented using lump-sum transfers (or taxes if negative) to households. Note that the parameter restrictions in (5) rule out the possibility that the central bank can increase consumption by printing money. Moreover, any profits from the purchase of capital K_0^{CB} are distributed lump-sum to households as well.[15] Thus, nominal transfers T_2 to households (or taxes, if $T_2 < 0$) are:

$$T_2 = \underbrace{K_0^{CB}\left(A_2\overline{P}_2 + A_1 P_1\right)}_{\substack{\text{output produced by capital} \\ \text{evaluated in nominal terms}}} + \underbrace{M_2^S - M_0^S}_{\substack{\text{change money supply} \\ \text{at } t=2}}. \tag{19}$$

The last term in Equation (19), $M_2^S - M_0^S$, denotes the change of the money supply at $t = 2$. For instance, if $M_0^S > \overline{M}$ and $M_2^S = \overline{M}$, the monetary injection at $t = 0$ is temporary, and thus the central bank taxes households at $t = 2$ to reduce the money supply to the initial level $\overline{M}$ (recall that households are endowed with $\overline{M}$ units of money at $t = 0$). If $M_0^S = \overline{M}$ and $M_2^S > \overline{M}$, the central bank is just intervening at $t = 2$. If $M_0^S = M_2^S > \overline{M}$, the monetary injection implemented at $t = 0$ is permanent.

4.3 Market-clearing conditions

The market-clearing conditions at $t = 0$ for capital, money, and deposits are:

$$\int K_0^b db + \int K_0^h dh + K_0^{CB} = \overline{K}, \int M_0^b db + \int M_0^h dh = M_0^S, \int D_0^b db = \int D_0^h dh. \tag{20}$$

If there is no monetary policy intervention, the amount of assets bought by the central bank in (20), K_0^{CB}, is zero, and the money supply is $M_0^S = \overline{M}$. The market clearing condition in the goods market at $t = 1$ is:

$$\int C_1^h dh = A_1 \overline{K}. \tag{21}$$

[15] Since the central bank is a large player in the market, I assume that the idiosyncratic shocks to K_0^{CB} cancel out. Thus, the overall stock of capital K_0^{CB} held by the central bank is unchanged at $t = 1$ and $t = 2$.

4.4 Equilibrium definition

The notion of equilibrium is similar to the one used in Section 3 and is standard. Given a monetary policy $\left\{M_0^S, M_2^S\right\}$, an equilibrium is a collection of prices $\left\{Q_0, P_1, r_2^K(\psi)\right\}$, households' choices $\left\{M_0^h, K_0^h, D_0^h, W_1^h, C_1^h, C_2^h\right\}$, banks' choices and return on deposits $\left\{M_0^b, K_0^b, D_0^b, r_2^b\left(\psi^b\right)\right\}$, and central bank's asset purchases K_0^{CB} and profits T_2 such that:

- households maximize utility;
- banks serve withdrawals at $t = 1$ until they run out of money (that is, if withdrawals W_1^h are constrained at zero for some households, Equation (14) must hold with equality); the return on deposits not withdrawn, $r_2^b\left(\psi^b\right)$, is paid using all the assets available to the bank at $t = 2$;
- the market-clearing conditions, (20) and (21), and the budget constraint of the central bank, (18), hold.

I consider symmetric equilibria in which all banks have the same deposits at $t = 0$.

5 Baseline model with banks: results

This section presents the results of the baseline model with piecewise-linear preferences in which banks offer deposits to households. The key results are obtained in Section 5.2 in an economy in which sufficiently many banks are hit by the negative shock to capital, ψ^L. Such banks become insolvent and are subject to runs at $t = 1$; as a result, households fly to money and away from deposits at $t = 0$ in anticipation of runs.

However, to clarify the result of the economy with runs, I first analyze a benchmark economy without bank runs in Section 5.1. To preclude runs, I shut down the negative shock to capital, ψ^L, by setting $\alpha = 0$.

The economy without bank runs provides a benchmark for the analysis of the economy with runs. In particular, some results of the economy with runs can be understood as an intermediate case between the bankless economy of Section 3 and the economy with no runs of Section 5.1.

5.1 Economy with no runs

I start by analyzing an economy in which I shut down the idiosyncratic shocks to capital in order to obtain an equilibrium without bank runs. This economy provides a benchmark against which the results of the economy with runs can be compared. To shut down the idiosyncratic shocks to capital, I set $\alpha = 0$. In this case, no agent is hit by the negative shock, ψ^L, and Equation (3) implies that the positive shock is $\psi^H = 0$. That is, it is as if shocks to capital did not happen.

The logic of this equilibrium is similar to the good equilibrium of Diamond and Dybvig (1983). All banks remain solvent at $t = 1$ and pay a positive return at $t = 2$, $r_2^b \geq 0$.[16] Banks offer deposits to households, which in turn withdraw at $t = 1$ only if they are hit by the "impatient" preference shock. Patient households are better off by not running and waiting until $t = 2$. The next proposition formalizes these results, and Appendix B.1 presents the proof.

Proposition 5.1. *(Economy with banks and no shocks to capital) Fix the money supply* $M_0^S = M_2^S = \overline{M}$. *If* $\alpha = 0$, *there exists an equilibrium with no runs and*

- *prices*

$$Q_0 = \frac{\beta}{1-\beta}\frac{\overline{M}}{\overline{K}} \equiv Q^*, \qquad P_1 = \frac{\overline{M}}{A_1\overline{K}} \equiv P^*\,; \tag{22}$$

- $t = 0$*: deposits* $D_0^h = D_0^b = D^*$, *where*

$$D^* \equiv \overline{M}/\kappa; \tag{23}$$

  money holdings $M_0^h = 0$ *and* $M_0^b = \kappa D^* = \overline{M}$*; and capital holdings* $K_0^h = \overline{K}\,(\beta + \kappa - 1)/\kappa\beta$ *and* $K_0^b = \overline{K}\,(1-\kappa)\,(1-\beta)\,/\,(\kappa\beta)$*;*
- $t = 1$*: withdrawals and consumption* $\left(W_1^h, C_1^h\right) = \left(D^*, \overline{C}\right)$ *if* h *is impatient, and* $\left(W_1^h, C_1^h\right) = (0, 0)$ *if* h *is patient;*
- $t = 2$*: return on capital* $1 + r_2^K = 1/\beta$, *return on deposits not withdrawn* $1 + r_2^b = 1/\beta$ *for all* b, *and consumption* $C_2^h = A_1\overline{K}\,(\beta - 1 + \kappa)\,/\kappa\beta$ *if* h *is impatient and* $C_2^h = A_1\overline{K}/\left[\beta\,(1-\beta)\right]$ *if* h *is patient.*

As in the good equilibrium of Diamond and Dybvig (1983), banks provide insurance against preference shocks, allowing impatient households to withdraw money and consume at $t = 1$ and patient households to receive a return on deposits at $t = 2$. Therefore, households hold no money ($M_0^h = 0$) at $t = 0$, because of the positive opportunity cost represented by the return on capital. Households prefer to hold banks' deposits, which have the advantages of both money and capital. That is, deposits can be withdrawn at $t = 1$ with certainty, and if not withdrawn, they pay the same return as capital at $t = 2$.

[16] Since there are no shocks, in this section I suppress the argument ψ in the notation of the return on capital and of the return on deposits, denoting them as r_t^K and r_2^b, respectively.

Given the price level P^*, D^* is the amount of deposits required to finance a household's consumption expenditure at $t = 1$ if the household is impatient. That is, at $t = 0$, households deposit all their endowment of money and a part of their endowment of capital into banks (in exchange for a promise to be able to withdraw money at $t = 1$ or to be repaid at $t = 2$) and invest the rest of their wealth into capital.[17]

The expected return on capital equals $1/\beta$; equivalently, the discounted return equals one. Given consumption $C_1^h = A_1\overline{K}/\kappa$ for impatient households, their marginal utility at $t = 1$ also equals one; see Equation (6). Thus, the marginal utilities of impatient households at $t = 1$ and $t = 2$ are equalized.

Banks invest a fraction κ of deposits into money in order to serve withdrawals by the fraction κ of impatient households at $t = 1$. The remaining fraction of deposits, $1 - \kappa$, is invested in capital. At $t = 2$, the return on capital is used to pay the return on deposits not withdrawn at $t = 1$.

Similar to the bankless economy, the price of consumption goods, P_1, is determined by equating consumption expenditures, $\int P_1 C_1^h dh$, to total money spent. The consumption expenditure can be rewritten as $P_1\, A_1\overline{K}$ using the market-clearing condition for goods. Unlike the bankless economy, here the entire money supply $\overline{M}$ is spent. This follows from the fact that banks hold the entire money supply at $t = 0$ ($M_0^b = \overline{M}$) and that all money withdrawn at $t = 1$ is spent. As a result, $P_1 = \overline{M}/\left(A_1\overline{K}\right)$.

Bankless economy and economy with no runs: a comparison. I now compare the price level and money velocity in the economy with no bank runs (Proposition 5.1) with the bankless economy of Section 3. In comparison to the bankless economy, banks offset the precautionary demand for money that arises in the bankless economy, reducing the demand for money and thus its equilibrium value. As a result, the price level P_1 is higher in the economy with banks and no runs, in comparison to the bankless economy; the next corollary summarizes this result.[18]

[17] Since the return on deposits not withdrawn equals the return on capital and there are no runs, any allocation with $D_t^h \in \left[D^*, \overline{M} + Q^*\overline{K}\right]$ corresponds to an equilibrium of this economy as well. That is, in comparison to the equilibrium of Proposition 5.1 in which $D_t^h = D^*$, households are indifferent between investing any extra dollar of wealth directly into capital or depositing it and letting banks invest on their behalf. However, if there were intermediation costs, households would be better off by holding only the minimum amount of deposits required to finance consumption at $t = 1$. The result $D_t^h = D^*$ can thus be viewed as arising from a limiting economy in which intermediation costs approach zero.

[18] A similar result arises in the monetary models of Brunnermeier and Sannikov (2011), Carapella (2012), and Cooper and Corbae (2002).

Corollary 5.2. *The price level is lower in the equilibrium of the bankless economy, in comparison to the economy with no runs. That is,* $(P_1 \text{ in bankless equilibrium}) < P^*$*, where* P^* *is the price level in the economy with no runs, defined by* (22)*.*

This result follows from the monetary nature of the model. To explain further, let v denote money velocity, defined implicitly by the equation of exchange:

$$M_0^S \times v = P_1 \left(A_1 \overline{K}\right). \tag{24}$$

Money velocity is endogenous and differs between the two equilibria. Less money is used for transactions in the bankless equilibrium than in the economy with no runs. As a result, velocity is lower in the bankless equilibrium, and thus the price level is lower as well.[19]

5.2 Economy with runs

This section presents the first main result of the paper. I analyze an economy in which many banks are hit by the negative shock to capital ψ^L, so that these banks become insolvent and are subject to runs at $t = 1$. In comparison to the economy with no bank runs, the equilibrium displays deflation (i.e., P_1 is lower) and a flight to liquidity (i.e., households increase money holdings, M_0^h, and reduce deposits, D_0^h).

The economy with runs can be understood as an intermediate case between the bankless economy of Section 3 and the economy with banks but no runs of Section 5.1. Consider money holdings M_0^h first; M_0^h are at an intermediate level between the two cases, $0 < M_0^h < \overline{M}$. In the bankless economy, there are no banks, and thus households hold the entire money supply in order to self-insure against preference shocks, $M_0^h = \overline{M}$. In the economy with banks but no runs, banks offer full insurance against preference shocks by providing deposits, and thus households hold no money, $M_0^h = 0$. In the intermediate case – the economy with runs – the insurance offered by banks is only partial because households at the end of the line are unable to withdraw. As a result, their money holdings are at an intermediate level between the two extreme scenarios. The same logic applies to

[19] The price of capital Q_0 in the bankless equilibrium is lower than Q^* as well, where Q^* is defined in Proposition 5.1. In the bankless economy, households' preference shocks are not insured; thus, the illiquidity of capital (i.e., its inability to provide insurance against preference shocks) reduces its demand and thus its price. In contrast, in the economy with banks and no runs (Proposition 5.1), banks provide sufficient insurance against preference shocks. In this case, the illiquidity of capital is irrelevant to households, and its nominal price Q_0 must be higher to clear the market.

deposits, and thus $0 < D_0^h < D^*$.[20] Nominal prices, Q_0 and P_1, and money velocity, v, are at an intermediate level as well, by a logic similar to Corollary 5.2.

Bank runs are a direct result of the idiosyncratic shocks to capital. Recall that banks invest in money and capital at $t = 0$, in amounts M_0^b and K_0^b, respectively. If bank b is hit by $\psi^L = -1$, the bank loses its stock of capital and thus becomes fundamentally insolvent. As a result, the return on deposits is negative: $r_2^b\left(\psi^L\right) = -1 < 0$; see (16). From the depositors' point of view, running on this bank is optimal because the return on money withdrawn is zero whereas the return on deposits not withdrawn is $r_2^b\left(\psi^L\right) < 0$.

The fact that runs are driven by fundamentals is similar to Allen and Gale (1998) but with an important difference. Unlike Allen and Gale (1998), only a fraction of banks are hit by the negative shock ψ^L and subject to a run at $t = 1$. Thus, the equilibrium of this section always involves runs on some banks at $t = 1$.[21] The identity of banks subject to runs is not known at $t = 0$ because the shocks to capital happen at $t = 1$. A further difference from Allen and Gale (1998) is that households in my model fly to money and reduce their holdings of deposits at $t = 0$ if they expect runs at $t = 1$. This difference is driven by the assumptions related to withdrawals introduced in Section 4.1 and discussed in Appendix A. The flight to money allows households to (partially) self-insure against preference shocks in the event of a run on their own bank at $t = 1$.

The next proposition summarizes the results of the economy with runs. Appendix B.2 provides the full specifications for equilibrium prices and allocations as a function of the parameters, presents the proof, and includes a thorough description of the maximization problem faced by households.

Proposition 5.3. *Fix the money supply $M_0^S = M_2^S = \overline{M}$. Under some parameter restrictions, there exists an equilibrium characterized by:*

- *prices $P_1 < P^*$ and $Q_0 < Q^*$;*
- *$t = 0$: households fly to liquidity, holding $M_0^h > 0$ and $0 < D_0^h < D^*$; banks invest a fraction κ of deposits into money, $M_0^b = \kappa D_0^b$, and the remainder in capital, $K_0^b = (1 - \kappa)\, D_0^b / Q_0$;*
- *$t = 1$: banks hit by ψ^L are subject to runs (i.e., both patient and impatient households want to withdraw at $t = 1$); impatient households holding deposits at banks not*

[20] In comparison to the economy with no runs of Section 5.1, households reduce D_0^h by depositing less money and a smaller fraction of their endowment of capital at banks, at $t = 0$.

[21] I can extend the model to allow two aggregate states at $t = 1$: one state in which idiosyncratic shocks ψ are realized and some banks are subject to runs, and another state in which idiosyncratic shocks are not realized and no bank is subject to runs. However, the main results would be unchanged.

subject to runs and impatient households at the beginning of the line at banks subject to runs consume $C_1^h > \overline{C}$; impatient households at the end of the line in a run consume $C_1^h < \overline{C}$;

- $t = 2$*: returns on deposits are* $r_2^b\left(\psi^H\right) > 0$ *and* $r_2^b\left(\psi^L\right) = -1$.

The results of Proposition 5.3 hold under the parameter restrictions provided in Appendix B.2. The key restriction is that α must be neither too large nor too small.[22] If α were too small, the probability of facing a run $t = 1$ would be low; therefore, the gains from flying to liquidity would be too small in comparison to the opportunity cost of holding money, and so households would hold no money, $M_0^h = 0$. If instead α were too large, households would be better off by holding no deposits at all, $D_0^h = 0$.

I conclude this section by explaining the result about the deposits contract that is offered by banks. As discussed in Section 4.1, the only relevant choice of banks is the fraction of deposits to be invested in money at $t = 0$. In equilibrium, banks invest a fraction κ of their deposits into money, as in the economy with no runs. First, note that to serve impatient households when a bank is not subject to a run, the bank must choose $M_0^b \geq \kappa D_0^b$. Second, I verify that households would be worse off by depositing in a bank that chooses $M_0^b > \kappa D_0^b$. By holding $M_0^b > \kappa D_0^b$, a bank would be able to serve more depositors that withdraw at $t = 1$ in the event of a run, but it would pay a lower return on deposits if not subject to a run. The welfare gains of the former effect do not offset the losses of the latter, because some of the depositors that withdraw in the event of a run and are patient and thus do not need to consume at $t = 1$. By a similar logic, a run-proof contract is not optimal either.[23]

5.3 Monetary injections

This section presents the second main result of the paper. In what follows, I analyze a temporary monetary injection implemented in the economy with runs (i.e., in the economy of Section 5.2). Recall that the initial endowment of money is $\overline{M}$; a temporary monetary injection is implemented by increasing the money supply at $t = 0$ to $M_0^S > \overline{M}$, keeping it constant at $t = 1$ at $M_1^S = M_0^S$, and reverting it back to $M_2^S = \overline{M}$ at $t = 2$. The key result

[22] Recall that α is the fraction of banks hit by ψ^L and subject to runs, and thus, by the law of large numbers, the probability that an household faces a run on its bank.

[23] A run-proof contract requires banks to invest 100% of their deposits into money; for an household, this is equivalent to holding no deposits and investing directly in money. Even though an equilibrium in which all households hold no deposits achieves the first best in bankless economy, a deviation by one single household to $D_0^h = 0$ is not optimal in the economy with runs because the existence of banks used by other households affect equilibrium prices and the opportunity cost of holding money.

is that such a temporary monetary injection does not affect prices (Q_0 and P_1 are constant), whereas it reduces money velocity and deposits; that is, a temporary injection amplifies the flight to liquidity. I provide intuition for this result in two ways. First, I decompose a temporary injection into a permanent injection at $t = 1$ and an anticipated, future reduction of money at $t = 2$. Second, I appeal to the mathematical structure of the model. Finally, I show that these results differ from those of the economy without runs, in which a temporary monetary injection does not affect either money velocity or deposits.

The next proposition formalizes the effects of a temporary monetary injection. I express all the results in terms of elasticities. For instance, the elasticity of deposits, D_0^h, with respect to a temporary monetary injection is given by $\left(dD_0^h/dM_0^S\right) \times \left(M_0^S/D_0^h\right)$; the elasticities of the other endogenous variables are defined similarly.

Proposition 5.4. *(Temporary monetary injection) Assume that parameters satisfy* (33)*. In an equilibrium with bank runs:*

$$\frac{dQ_0}{dM_0^S} \times \frac{M_0^S}{Q_0} = 0, \qquad \frac{dP_1}{dM_0^S} \times \frac{M_0^S}{P_1} = 0, \qquad \frac{d\overline{P}_2}{dM_0^S} \times \frac{M_0^S}{\overline{P}_2} = 0$$

$$\frac{dv}{dM_0^S} \times \frac{M_0^S}{v} = -1, \qquad \frac{dD_0^h}{dM_0^S} \times \frac{M_0^S}{D_0^h} < 0, \qquad \frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} > 1.$$

These results are related to the endogenous determination of money velocity v, defined in Equation (24). If v were constant, an increase in M_0^S would trigger an increase in nominal prices Q_0 and P_1. However, money velocity in the model is *endogenous* and drops as a result of the monetary injection. In this baseline model, the endogenous reduction in velocity offsets the direct effect of the monetary injections, and thus prices are unchanged. Note that money held by households, M_0^h, increases more than one-for-one with the monetary injection. That is, households not only absorb all the monetary injection of the central bank but also keep in their wallets some of the money that, absent the injection, they would have deposited at banks.

The result of Proposition 5.4 can be understood by decomposing a temporary monetary injection into two separate interventions: a permanent injection implemented at $t = 0$ and a reduction of money supply implemented at $t = 2$ but announced at $t = 0$. That is, a temporary injection is the "sum" of these two interventions. I now analyze these two interventions separately, and then comment further on the fact that households change D_0^h and M_0^h in response to a monetary despite the prices that they face are unchanged.

A permanent injection at $t = 0$ implies that money supply increases at $t = 0$ to $M_0^S > \overline{M}$ and stays constant afterward, $M_1^S = M_0^S$ and $M_2^S = M_0^S$. The effects of this intervention on velocity and prices are standard. Money velocity remains constant, and thus prices respond one-for-one with the monetary injection; more precisely, the elasticity of nominal prices with respect to the monetary injection is one. Using the assumption about the "price level after the crisis" $\overline{P}_2$, in Equation (5), $\overline{P}_2$ responds one-for-one with the injection as well. Since the effects of this permanent injection are purely nominal, the real quantities of money and deposits held by households, M_0^h/P_1 and D_0^h/P_1, are unchanged. However, nominal money and nominal deposits, M_0^h and D_0^h, must respond one-for-one as well (i.e., their elasticity with respect to the monetary injection is one) because the price level, P_1, is higher. The next proposition summarizes this result.

Proposition 5.5. *(Permanent monetary injection, $t = 0$) Let $M_0^S = M_1^S = M_2^S = \widehat{M}$, so that any change to the money supply implemented at $t = 0$ is permanent. In an equilibrium with bank runs:*

$$\frac{dQ_0}{d\widehat{M}} \times \frac{\widehat{M}}{Q_0} = 1, \qquad \frac{dP_1}{d\widehat{M}} \times \frac{\widehat{M}}{P_1} = 1, \qquad \frac{d\overline{P}_2}{d\widehat{M}} \times \frac{\widehat{M}}{\overline{P}_2} = 1$$

$$\frac{dv}{d\widehat{M}} \times \frac{\widehat{M}}{v} = 0, \qquad \frac{dD_0^h}{d\widehat{M}} \times \frac{\widehat{M}}{D_0^h} = 1, \qquad \frac{dM_0^h}{d\widehat{M}} \times \frac{\widehat{M}}{M_0^h} = 1.$$

I now turn to the analysis of a reduction in the money supply at $t = 2$. This intervention triggers a one-for-one reduction in $\overline{P}_2$; see (5). Since the intervention at $t = 2$ is anticipated at $t = 0$, it affects Q_0 and P_1, too. In particular, I claim that Q_0 and P_1 must respond one-for-one with a change in $\overline{P}_2$, and thus, one-for-one with the monetary injection. The one-for-one link between Q_0 and P_1, on the one hand, and $\overline{P}_2$, on the other, is a by-product of the local linearity of households' utility. To clarify the role of the utility function, let me focus on households' money demand, although a similar logic applies to households' demand for deposits and capital.

From a partial equilibrium perspective, the local linearity of households' utility implies a very large elasticity of money demand with respect to its price. Such price is given by the opportunity cost of holding money, represented by the expectation of the nominal return on capital defined in Equation (4): $\mathbb{E}\left\{1 + r_2^K(\psi)\right\} = \left(\overline{P}_2 A_2 + P_1 A_1\right)/Q_0$.[24] If a monetary injection triggered a change in the opportunity cost of holding money, households' money

[24]The expectation is taken with respect to the shock to capital, ψ.

demand would change dramatically because of the large elasticity, thereby violating market clearing in the money market. To obtain market clearing in the money market, Q_0 and P_1 must adjust one-for-one with $\overline{P}_2$ so that the opportunity cost of holding money $\mathbb{E}\left\{r_2^K(\psi)\right\}$ is unchanged.

Next, I claim that a reduction in money supply at $t = 2$ also amplifies the flight to liquidity at $t = 0$ (i.e., it reduces deposits, D_0^h, and increases money held by households, M_0^h). A contraction in the money supply at $t = 2$ reduces P_1, as explained before. Moreover, the equation of exchange, (24), implies that P_1 is determined by the quantity of money in circulation at $t = 0$ (recall that $M_0^S = M_1^S$, as assumed in Section 4.2) and by money velocity. Money at $t = 0$ is unchanged at $M_0^S = \overline{M}$ because the monetary injection happens at $t = 2$; therefore, a drop in P_1 requires a drop in velocity. A lower velocity is obtained if the flight to liquidity is amplified. The negative link between money velocity and the flight to liquidity is exemplified by the fact that velocity is low in the bankless equilibrium, in which the flight to liquidity is maximal ($D_0^h = 0$ and $M_0^h = \overline{M}$; i.e., households hold the entire money supply), and high in the economy with banks and no runs, in which there is no flight to liquidity ($D_0^h = D^*$ and $M_0^h = 0$). The next proposition summarizes these results.[25]

Proposition 5.6. *(Monetary injection at $t = 2$) Assume that parameters satisfy (33) and that any change to the money supply at $t = 2$ is anticipated at $t = 0$. Then, in an equilibrium with bank runs:*

$$\frac{dQ_0}{dM_2^S} \times \frac{M_2^S}{Q_0} = 1, \qquad \frac{dP_1}{dM_2^S} \times \frac{M_2^S}{P_1} = 1, \qquad \frac{d\overline{P}_2}{dM_2^S} \times \frac{M_2^S}{\overline{P}_2} = 1$$

$$\frac{dv}{dM_2^S} \times \frac{M_2^S}{v} = 1, \qquad \frac{dD_0^h}{dM_2^S} \times \frac{M_2^S}{D_0^h} > 0, \qquad \frac{dM_0^h}{dM_2^S} \times \frac{M_2^S}{D_0^h} < 0.$$

An alternative explanation of Proposition 5.4 is based on the mathematical structure of the model. Note that households change D_0^h and M_0^h in response to a temporary monetary injection despite prices remaining unchanged. This is because the households' first-order conditions with respect to D_0^h and M_0^h – evaluated at the equilibrium – depend only on prices Q_0 and P_1. In other words, these first-order conditions are independent of D_0^h and M_0^h because of the local linearity of preferences – recall that a household's utility is linear

[25]Note that Proposition 5.6 presents the *elasticity* of endogenous variables with respect to an anticipated change of money at $t = 2$. To understand the effects of a reduction of money at $t = 2$, the signs of the results in Proposition 5.6 must be flipped.

in consumption C_1^h and C_2^h, which are, in turn, linear in D_0^h and M_0^h using (15) and (17). Moreover, the first-order conditions are independent of the money supply M_0^S too, because the central bank's decisions do not enter directly into the households' utility maximization problem. As a result, the price of capital Q_0 and the price level P_1 that sustain the two first-order conditions do not depend on M_0^S, and thus a temporary monetary injection does not change Q_0 or P_1.

This mathematical structure provides guidance for the intuition behind the results. Temporary monetary injections that vary M_0^S change deposits D_0^h and money holdings M_0^h even though, in this simple model, they do not change prices. It is not entirely correct to explain the effects of a temporary injection by appealing to the impact that such policy has on prices. Thus, the best and most general way to provide intuition for Proposition 5.4 is to decompose a temporary injection into a permanent injection at $t = 1$ and a permanent, anticipated contraction at $t = 2$. As I discuss in Section 6.3.2, this argument applies also to the model with standard preferences of Section 6.

Next, I analyze a temporary monetary injection in the economy with no bank runs (i.e., in the equilibrium of Proposition 5.1). The objective is to highlight the role of bank runs and of the flight to liquidity in shaping the effects of temporary injections. The next proposition shows that the effects of a temporary monetary injection if there are no runs are purely nominal. The price level, P_1, responds one-for-one to the temporary monetary injection, whereas real deposits D_0^h/P_1, velocity v, and households' money holdings M_0^h are not affected.

Proposition 5.7. *(Temporary injection in economy with no runs) If $\alpha = 0$ and $M_2^S = \overline{M}$, there exists an equilibrium with $M_0^h = 0$, $D_0^h > 0$, and no bank runs. Moreover:*

$$\frac{dP_1}{dM_0^S} \times \frac{M_0^S}{P_1} = 1, \quad \frac{dv}{dM_0^S} \times \frac{M_0^S}{v} = 0, \quad \frac{d\left(D_0^h/P_1\right)}{dM_0^S} \times \frac{M_0^S}{\left(D_0^h/P_1\right)} = 0, \quad \frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} = 0.$$

Similar to the economy with runs and Proposition 5.4, the results are the "sum" of the effects of a permanent injection at $t = 0$ and of an anticipated reduction in money at $t = 2$. However, in the economy without runs, the reduction in money at $t = 2$ does not change households' deposits and money holdings. Since there are no runs, households deposit all their money in banks, and thus a deflationary pressure does not change households' incentives to hold no money in their wallets, as long as this pressure is small enough so that the opportunity cost of holding money remains positive.

To sum up, Proposition 5.4 shows that (i) a temporary monetary injection reduces deposits, and (ii) the elasticity of P_1 with respect to a temporary monetary injection is zero. Section 6 presents a model with a more standard utility function; in this model, the effects of a temporary monetary injection on D_0^h and M_0^h are more subtle, and P_1 responds to a monetary injection as well. Nonetheless, quantitative examples that are calibrated to the Great Depression and the Great Recession show that the main message is unchanged and the unintended consequences of temporary injections are large.

6 Smooth-utility model

Using the baseline model with piecewise-linear utility, Equation (2), I have shown that temporary monetary injections amplify the flight to liquidity in an economy with bank runs. The local linearity implied by Equation (2) simplifies the policy analysis because households are indifferent among any quantity of money, deposits, and capital as long as their first-order conditions hold with equality. However, such linearity also implies a very high elasticity of money demand with respect to the opportunity cost of holding money, raising a possible concern about the policy analysis.

This section studies the robustness of the results using a variant of the baseline model with a more standard smooth utility function. That is, I replace the piecewise-linear utility function with log utility.

To study the economy with runs and temporary injections in this richer model, I have to rely on numerical methods. I present two quantitative examples calibrated to study the Great Depression and the run on money market mutual funds that took place in 2008, respectively. Different from the baseline model, nominal prices increase with a temporary monetary injection, although less than one-for-one because of the endogenous reduction in velocity. Nonetheless, the degree of monetary non-neutrality is large. That is, the elasticities of the key endogenous variables with respect to a temporary monetary injection are approximately the same as those in Proposition 5.4.

In addition, the calibrated model allows me to disentangle the two sources that affect the flight away from deposits and toward money: the increase in the probability of runs and the monetary injections. In the analysis of the Great Recession, I run a counterfactual policy analysis in which I shut down the monetary injections implemented by the Federal Reserve. By comparing the results with and without monetary injections, I quantify the contribution of the Federal Reserve's intervention to the flight to liquidity.

Finally, I close the section with a brief comment on welfare and optimal policy.

6.1 Preferences

This section presents the preferences of the smooth-utility model that is used to perform the quantitative exercises. The rest of the model is the same as in Sections 2 and 4.

I replace the utility function in Equation (1) with

$$\text{utility} = \mathbb{E}_0 \left\{ \varepsilon_1^h \, u\left(C_1^h\right) \right\} + \beta C_2^h, \tag{25}$$

where ε_1^h is a preference shock realized at $t = 1$ and taking values:

$$\varepsilon_1^h = \begin{cases} \varepsilon^H & \text{(impatient household)} \quad \text{with probability } \kappa \\ \varepsilon^L & \text{(patient household)} \quad \text{with probability } 1-\kappa \end{cases}, \varepsilon^H > 1 > \varepsilon^L \geq 0. \tag{26}$$

Preference shocks are now represented by ε_1^h, whose realization is private information, similar to the baseline model.[26] I allow for a general formulation in which ε^L can be positive so that even patient households may want to consume at $t = 1$. This feature opens up a role for banks to increase welfare in comparison to a bankless economy (see Section 6.4) and is important in the quantitative examples. Without loss of generality, I impose the normalization $\mathbb{E}_0 \left\{ \varepsilon_1^h \right\} = 1$.

I use the functional form $u(C) = \overline{C} \log C$, where $\overline{C}$ is a constant. Under this parameterization, households' marginal utilities vary endogenously, creating a richer feedback between policy interventions and households' choices. Moreover, I replace the restriction on $\overline{C}$ in Equation (6) with

$$\overline{C} = A_1 \overline{K}. \tag{27}$$

The role of this restriction is similar to that in (6). When time-1 consumption of patient and impatient households is evaluated at the first best, Equation (27) implies that their marginal utilities equal one and are thus equal to the marginal utility of consumption at $t = 2$.

6.2 Economy with no bank runs

To study the economy with no runs, I set $\alpha = 0$ so that no bank is subject to the negative shock ψ^L. The logic of the equilibrium is the same as in Section 5.1.

The only difference in comparison to the equilibrium in Section 5.1 is that patient

[26] I also assume that ε_1^h is i.i.d. across households and that the law of large numbers holds, also similar to the baseline model.

households enjoy some utility from consumption if $\varepsilon^L > 0$. Nonetheless, patient households want to consume less than impatient households because $\varepsilon^L < \varepsilon^H$. Since withdrawals at $t = 1$ are restricted to be either zero or D_0^h (see Section 4.1), a household holds a positive amount of money at $t = 0$, $M_0^h > 0$, and it uses this money to finance consumption if $\varepsilon_1^h = \varepsilon^L$. That is, a patient household does not withdraw at $t = 1$ and consumes M_0^h / P_1. If instead $\varepsilon_1^h = \varepsilon^H$, household h finances its consumption expenditure using not only its money, M_0^h, but also withdrawals, W_1^h, and thus consumes $\left(M_0^h + W_1^h\right) / P_1$.

The next proposition formalizes the equilibrium in the economy with no runs.

Proposition 6.1. *If $M_0^S = M_2^S = \overline{M}$ and $\alpha = 0$, there exists an equilibrium with no runs:*

$$D_0^h = \overline{M}\frac{\varepsilon^H - 1}{1-\kappa}, \quad M_0^h = \overline{M}\frac{1-\kappa\varepsilon^H}{1-\kappa}, \quad M_0^b = \kappa D_0^b, \quad Q_0 = \frac{\beta}{1-\beta}\frac{\overline{M}}{\overline{K}},$$

$$C_1\left(\varepsilon^H\right) = A_1\overline{K}\varepsilon^H, \quad C_1\left(\varepsilon^L\right) = A_1\overline{K}\,\frac{1-\kappa\varepsilon^H}{1-\kappa}, \quad P_1 = \frac{\overline{M}}{A_1\overline{K}}, \quad r_2^K = r_2^b = \frac{1}{\beta} - 1,$$

where $C_1\left(\varepsilon^h\right)$ denotes the consumption of a household hit by the preference shock ε^h.

6.3 Economy with bank runs: quantitative examples

I now turn to the analysis of the equilibrium with bank runs. As in the baseline model, a positive and sufficiently large value of α implies that some banks become insolvent and subject to runs at $t = 1$, so that households fly to liquidity at $t = 0$. Because of the richer model, I cannot solve for the equilibrium in closed form, and therefore, I rely on numerical methods. Nonetheless, the logic of the result is identical to Section 5.2.

I provide two quantitative examples calibrated to the Great Depression and the Great Recession, respectively. The objective of these quantitative examples is twofold. First, I show that the results of the model with piecewise-linear utility (Sections 2-4) are robust to the model with more standard preferences. Second, I focus on actual crisis episodes – the Great Depression and the Great Recession – to quantify the force that gives rise to the unintended consequences of temporary monetary injections. Let me emphasize that the model is very simple and abstracts from features that might be relevant in practice and in richer model. Nonetheless, while other forces might be at work in practice and in a richer model, my simple framework allows me to shut down all such forces in order to provide some evidence about the magnitude of the channel that I have identified.

6.3.1 Comparison with the Great Depression

I develop the quantitative example applied to the Great Depression in three steps. I first calibrate the model, choosing most of the parameters to match the equilibrium with no bank runs (Proposition 6.1) with the data in 1929 (i.e., before the first wave of bank runs) and to match some key facts of the Great Depression. All the data are from Friedman and Schwartz (1963, 1970), except those about the price level, which are from the NBER Macrohistory Database. Second, I solve for the equilibrium with bank runs and compare deposits D_0^h, money holdings by households M_0^h, price level P_1, and money velocity v with their data counterparts in March 1933 (i.e., the peak of the Depression). The model accounts for an important fraction of the movements of these key macroeconomic and banking variables during the Depression. In the final step, I compute the elasticities of the main endogenous variables with respect to a temporary monetary injection, showing that the results are similar to those derived in the model with a piecewise-linear utility function.

Let me emphasize that studying temporary monetary injections is relevant for the analysis of the Great Depression. Friedman and Schwartz (1963) document that some of the monetary injections implemented during this crisis were temporary. For instance, the Federal Reserve increased credit during the first banking crisis (October 1930 to January 1931), but credit decreased as bank failures declined in early 1931. A similar contraction in the money supply took place in February and March 1932, when bank failures tapered off.

Calibration. To match the equilibrium with no runs to the 1929 data, I set the endowment of money to $\overline{M} = \$7.185$ billion (matching the money supply in 1929), the fraction of patient agents $\kappa = 0.077$, and the preference shocks parameter $\varepsilon^H = 6.48$; the value of ε^L is determined residually by the normalization $\mathbb{E}\left\{\varepsilon_t^h\right\} = 1$, implying $\varepsilon^L = 0.54$. In the economy with no runs (Proposition 6.1), these choices imply $M_0^h = \$3.9$ billion, $D_0^h = \$42.7$ billion, and $M_0^b = \$3.285$ billion, matching currency held by the public, total deposits, and money held by banks in 1929, respectively.

Next, I calibrate α, M_0^S, and M_2^S to match some key facts of the Great Depression. I set $\alpha = 0.1$ (i.e., 10% of the banks in the model are hit by the shock to capital $\psi^L = -1$); this value corresponds to the fraction of banks that suspended operations during the Great Depression, weighted by the volume of their deposits. I set $M_0^S = \$8.44$ billion, matching the 17.5% increase in the money supply between August 1929 and March 1933. I set $M_2^S = \$6.39$ billion (i.e., the money supply in the model drops in the years after 1933), so that $\overline{P}_2$ in the economy with runs is 11% lower than the price level in the equilibrium with

Table 1: Comparison with the Great Depression

Variable	Economy no runs (and 1929 data)	Economy with runs (difference from economy with no runs)	Data (difference 1933-1929)
Deposits $\left(D_0^h\right)$	42.7	-6.68	-17
Money, banks $\left(M_0^b\right)$	3.285	-0.51	-0.185
Money, households $\left(M_0^h\right)$	3.9	+1.77	+1.67
Money velocity (v)	1	-0.25	-0.33
Price level (P_1)	100	-11.4	-33.3%

Parameter values: $\beta = 0.94$, $\kappa = 0.077$, $\alpha = 0.1$, $\overline{M} = 7.185$, $\overline{K} = 1.1257$, $A_1 = 0.0638$, $M_0^S = \$8.44$, $M_2^S = \$6.39$. Price level data are from the NBER Macrohistory Database; other data are from Friedman and Schwartz (1963, 1970). Deposits and money are in billions of dollars.

no runs, P^*, matching the 11% drop in the price level between 1929 and 1937.[27]

Finally, I normalize A_1 and $\overline{K}$ so that prices in the equilibrium with no runs are $Q^* = 100$ and $P^* = 100$, and I set the discount factor β to 0.94. This low value of β is consistent with the three-year period between the acute phase of the crisis in 1933 (periods $t = 0, 1$ in the model) and the peak of the recovery in 1937 (period $t = 2$ in the model).

Results. Table 1 presents the results. By construction, the results of the economy with no runs match the 1929 data. In the economy with runs, deposits drop by \$6.68 billion, whereas deposits dropped by \$17 billion in the data (from \$42.7 in 1929 to \$25.7 billion in 1933). Thus, the model accounts for almost 40% of the drop in deposits. Although the model underestimates the drop in deposits in the Great Recession, it overestimates the drop in money held by banks (-\$0.51 billion in the model vs. -\$0.185 billion in the data). This is because, in the model, banks' reserves are the same in economy with and without runs (money holdings by banks is $M_0^b = \kappa D_0^b$ in the economies both with and without runs, and thus banks' reserves are a constant fraction κ of deposits), whereas banks' reserves increased in the data (from 7.7% in 1929 to 11.8% in 1933). That is, the data show a flight to liquidity by banks that the model is not able to replicate. Since the model overestimates the drop in money held by banks, it mechanically overestimates the flight to liquidity by households because money is held by either banks or households. Finally, the

[27] In the model, the only way to obtain $\overline{P}_2 < P^*$ is to have a contraction in the money supply at $t = 2$ (i.e., $M_2^S < \overline{M}$). That is, the model focuses on the events of the acute phase of the Depression and abstracts from the factors that made the recovery slow, requiring $M_2^S < \overline{M}$ to replicate the low price level in 1937.

model accounts for about one-third of the drop in prices observed in the data and two-thirds of the drop in money velocity.

Finally, I study the effects of temporary injections by computing the same elasticities studied in Proposition 5.4 for the baseline model. That is, I compute the elasticity of velocity v, price level P_1, deposits D_0^h, and money held by households M_0^h with respect to a change in M_0^S, and evaluate such elasticities at the equilibrium matching the Depression:[28]

$$\frac{dv}{dM_0^S} \times \frac{M_0^S}{v} = -0.93, \qquad \frac{dP_1}{dM_0^S} \times \frac{M_0^S}{P_1} = 0.05,$$

$$\frac{dD_0^h}{dM_0^S} \times \frac{M_0^S}{D_0^h} = -0.13, \qquad \frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} = 1.55.$$

The results are similar to those of Proposition 5.4; that is, the degree of monetary non-neutrality is very high.

6.3.2 Comparison with the Great Recession

I now present a quantitative example calibrated to the Great Recession and, in particular, to the run on money market mutual funds (MMMFs) that took place in September 2008. I first provide a brief background and explain why the model is relevant to analyzing this event, and then present calibration and results. Similar to the experiment in Section 6.3.1, I compare the days before the Lehman Brothers collapse (before September 15, 2008) to the economy without runs, the month after the Lehman collapse to $t = 0, 1$ in economy with runs, and the following months to $t = 2$ in the economy with runs.

Appendix C provides further discussion about the comparison of the model with the data. The Appendix also shows that the results are robust to possible concerns related to the timing of the monetary injections, the effects of the guarantee program extended by the Treasury on MMMFs, and the calibration of preference shocks.

The run on money market mutual funds in September 2008. Duygan-Bump et al. (2013) and Schmidt, Timmermann, and Wermers (2016) document a large run on prime institutional MMMFs in the month that followed the bankruptcy of Lehman Brothers on September 15, 2008.[29] These funds, which managed about \$1.3 trillion before September

[28] I compute the elasticity with respect to v by approximating dv/dM_0^S with $\Delta v/\Delta M_0^S$, where $\Delta M_0^S = 0.01 \times M_0^S$ and Δv is the difference between the velocity computed when $M_0^S = \$8.44 \times (1 + 1\%) = \$8.52$ billion and the velocity computed when $M_0^S = \$8.44$ billion. I use a similar approach for the other elasticities.

[29] Prime institutional MMMFs are marketed to institutional investors and invest mostly in instruments other

15, experienced $400 billion in redemptions during the run. On September 19, the Federal Reserve announced a new facility with the objective of providing liquidity to MMMFs, the Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility (AMLF). Since the Federal Reserve could not lend directly to MMMFs, it designed the AMLF so as to provide funding indirectly. The AMLF extended nonrecourse loans to "traditional" banks; these loans were collateralized by the asset-backed commercial paper that traditional banks purchased from MMMFs, which in turn used the proceeds of the sales to pay redemptions.

Before turning to the quantitative analysis, I argue that the model is relevant to analyzing this event. I do so in three steps. First, I suggest a reinterpretation of the model in order to apply it to MMMFs. Second, I discuss the role of liquidity in the model and relate it to MMMFs. Third, I explain how the model can produce an equilibrium with runs but no deflation, in order to match a key fact of the 2008 data.

When applying the model to MMMFs, households must be interpreted as investors of prime institutional MMMFs, deposits as shares of these MMMFs, capital as asset-backed commercial paper, and money as liquid assets such as M1 or US Treasuries.[30] I also claim that asset purchases by the central bank in the model can be used to model the AMLF. That is, the results are quantitatively identical if I add "traditional banks" to the model in order to implement the AMLF as the Federal Reserve did. In this case, traditional banks in the model would buy capital (interpreted as commercial paper) from banks (interpreted as MMMFs). Traditional banks' purchases of capital would be funded with fully collateralized loans from the central bank (interpreted as loans extended under AMLF). This extension would be consistent with Begenau, Bigio, and Majerovitz (2016), who provide evidence of a reallocation of assets from shadow banks (including MMMFs) to traditional banks with funding provided, at least in part, by the liquidity facilities of the Federal Reserve.

In terms of the liquidity services provided by MMMFs, the model would produce the same results if the demand for money arose because of money in the utility function rather than a cash-in-advance constraint.[31] The money in utility approach is a more general way

than Treasury securities and agency debt. An important fraction of their investments are in asset-backed commercial paper.

[30]See Krishnamurthy and Vissing-Jorgensen (2012) for evidence on the liquidity services provided by US Treasuries.

[31]This can be obtained with a utility of the form $C_1^h + \varepsilon_1^h V\left[\left(M_0^h + W_1^h\right)/P_1\right] + \beta C_2^h$, where V is a strictly increasing and strictly concave function. As in many other monetary models, modeling money using a cash-in-advance constraint or money in the utility function produces the same results for some analyses.

to motivate a demand for shadow banks liabilities (see, e.g., Nagel, 2016).

Finally, note that monetary injections can produce an equilibrium with runs and flight to liquidity but *no deflation*, as in 2008. In the model with log utility, monetary injections affect not only velocity and deposits but also prices. In particular, in the calibration presented below, monetary injections allow the economy with bank runs to achieve the same price level as the economy without runs (i.e., $P_t = P^*$ in the economy with runs and monetary injections), consistent with the 2008 data.

Calibration. I choose the value of κ, ε^H, and $\overline{M}$ to match the equilibrium in the economy with no runs to the data before September 15, 2008 (i.e., before the collapse of Lehman Brothers); the value of ε^L is then determined by the normalization $\mathbb{E}_0\left\{\varepsilon_1^h\right\} = 1$ introduced in Section 6.1. I set $\kappa = 0.2$, so that banks in the model hold 20% of deposits in money (recall that $M_0^b = \kappa D_0^b$, from Proposition 6.1); this is based on the data of Duygan-Bump et al. (2013) and Schmidt, Timmermann, and Wermers (2016), who report, respectively, that the MMMFs involved in the run held 21.3% and 18.57% of their portfolios in liquid assets. Next, I turn to ε^H; note that $1 < \varepsilon^H \leq 1/\kappa$, where the lower bound is imposed in Equation (25), whereas the upper bound follows from $\mathbb{E}_0\left\{\varepsilon_1^h\right\} = 1$ and $\varepsilon^L \geq 0$. Since I have no direct evidence to calibrate ε^H, I choose the value that produces the most conservative results about the policy analysis, which is given by the limit case $\varepsilon^H = 1/\kappa = 5$ (and thus $\varepsilon^L = 0$, using $\mathbb{E}_0\left\{\varepsilon_1^h\right\} = 1$). In Appendix C.3, I consider an alternative calibration of ε^L and ε^H that produces bigger unintended consequences of temporary injections. I set $\overline{M} = \$260$ billion, so that deposits in the economy with no runs are $D^* = \overline{M}/\kappa = \$1.3$ trillion (see Proposition 6.1). Since deposits in the model are equal to assets held by banks, this calibration matches the \$1.3 trillion assets held by MMMFs before September 15.

Next, I calibrate M_0^S, M_2^S, and α to match some key facts observed during the run on MMMFs. I set $M_0^S = \$410$ billion, implying that the central bank in the model injects $M_0^S - \overline{M} = \$150$ billion, in line with the monetary injection of the Federal Reserve implemented using the AMLF. I set $M_2^S = \overline{M}$, matching the fact that the AMLF was a temporary facility (it was closed on February 1, 2010); setting $M_2^S = \overline{M}$ also implies price stability after the crisis, $\overline{P}_2 = P^*$.[32] Finally, I choose α to match price stability (i.e., P_1 in the economy with runs and monetary injections equals P^*). The resulting value is very small,

[32]The temporary nature of the AMLF confirms once more the relevance of studying temporary monetary injection. Other monetary injections implemented by the Federal Reserve in 2008 are also better characterized as temporary (even though the balance sheet of the Federal Reserve did not shrink after the peak of the crisis) because the Federal Reserve started to pay interest on reserves.

$\alpha = 0.0011$; that is, only 0.11% of banks in the model are hit by ψ^L at time $t = 1$.

Finally, similar to the quantitative example in Section 6.3.1, I normalize A_1 and $\overline{K}$ so that prices in the economy with no runs are $Q^* = 100$ and $P^* = 100$, and I set the discount factor to $\beta = 0.998$. This choice of β is motivated by the short period of time under analysis, that is, one month.

Results. In the economy with runs, deposits are $D_0^h = \$1.114$ trillion. The difference between D_0^h and deposits in the economy without runs ($D^* = \$1.3$ trillion) represents redemptions from MMMFs in the model: $D_0^h - D^* = \$186$ billion.[33] Comparing this number with the \$400 billion redemptions in the data, I conclude that the model accounts for $\$186/\$400 = 46.5\%$ of the redemptions from prime MMMFs that took place after Lehman's collapse. The model also produces a flight to money of the same magnitude, $M_0^h = \$187$ billion in comparison to $M_0^h = 0$ in the economy without runs.

Next, I compute the elasticities of velocity v, price level P_1, deposits D_0^h, and money held by households M_0^h with respect to a change in money supply M_0^S at the equilibrium that replicates the run on MMMFs:[34]

$$\frac{dv}{dM_0^S} \times \frac{M_0^S}{v} = -0.99, \qquad \frac{dP_1}{dM_0^S} \times \frac{M_0^S}{P_1} = 0.004, \tag{28}$$

$$\frac{dD_0^h}{dM_0^S} \times \frac{M_0^S}{D_0^h} = -0.45, \qquad \frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} = 2.73. \tag{29}$$

Similar to the Great Depression, the degree of monetary non-neutrality of the model is very large because P_1 is almost unchanged in response to a change in M_0^S. That is, these elasticities are in line with the results of Proposition 5.4.

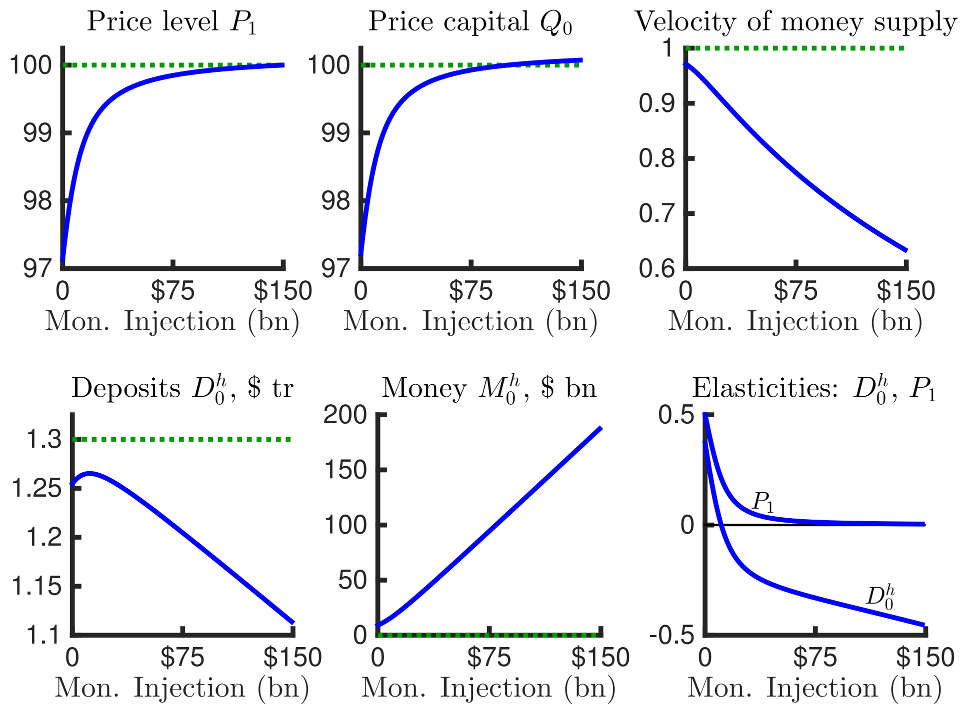
Finally, I ask what would have happened if the Federal Reserve had injected a smaller amount of money into the economy, providing results for any value of the monetary injection between \$0 and \$150 billion in Figure 3. To understand this last exercise, note that the elasticities with respect to M_0^S in (28) and (29) are *local*; that is, they describe the effects of a small temporary injection *evaluated at the equilibrium that replicates the 2008 run*. In contrast, the policy exercise in Figure 3 studies the *global* impact of monetary injections.

The top panels of Figure 3 depict the results that are related to prices and money velocity. Without monetary injections, velocity would have been higher, and the price level

[33]That is, I compare the runs in the data with $D_0^h - D^*$ in the model, rather than with the events at $t = 1$ in the model. See Appendix C.1 for further discussion of the runs at $t = 1$ in the model versus the data.

[34]I compute the elasticities using the same approach described in Section 6.3.1.

Figure 3: Monetary injections, Great Recession



The green dotted line represents the equilibrium with no runs and no monetary injections; the blue solid line represents the equilibrium with runs, as a function of the monetary injection. The monetary injection (horizontal axis) corresponds to the difference $M_0^S - \overline{M}$. The bottom right panel depicts $\left(dD_0^h/dM_0^S\right) \times \left(M_0^S/D_0^h\right)$ and $\left(dP_1/dM_0^S\right) \times \left(M_0^S/P_1\right)$. Parameter values: $\beta = 0.998$, $\kappa = 0.2$, $\alpha = 0.0011$, $\overline{M} = \$260$ billion, $\overline{K} = 1,297.4$, $A_1 = 0.002$, $M_2^S = \overline{M}$.

would have dropped 2.86% the month after the Lehman bankruptcy. That is, the monetary injection reduced velocity just as in the baseline model, but it also offset deflation. In the bottom of Figure 3, the left and middle panels depict deposits and money held by households, respectively. Without monetary injections, $D_0^b = \$1.255$ trillion, and thus redemptions would have been only \$45 billion. Similarly, the flight to money (M_0^h) would have been only \$9 billion.

The bottom right panel of Figure 3 depicts the elasticities of deposits and of the price level with respect to M_0^S; that is, $\left(dD_0^h/dM_0^S\right) \times \left(M_0^S/D_0^h\right)$ and $\left(dP_1/dM_0^S\right) \times \left(M_0^S/P_1\right)$. These results are consistent with the analysis of the baseline model with piecewise-linear preferences. The amplification of the flight to liquidity is maximal (i.e., the elasticity of D_0^h is more negative) in scenarios where P_1 changes the least. When the central bank does not inject any money into the economy, the elasticity of P_1 with respect to M_0^S is about 0.5

and the elasticity of deposits is positive; that is, temporary injections dampen the flight to liquidity. At the extreme (i.e., when the central bank injects \$150 billion), the elasticity of P_1 is 0.04 and the elasticity of deposits is -0.45; that is, the amplification of the flight to liquidity is maximal.

Moreover, the unintended effects of temporary injections are economically relevant. According to the model, the \$150 billion injected by the Federal Reserve using AMLF amplified the run on prime MMMFs by \$141 billion ($= \$186 - \$45$) and the flight to liquidity by \$178 billion ($= \$187 - \$9$).

6.4 Welfare and optimal policy: Discussion

I conclude this section with some remarks on welfare and optimal policy. I first discuss the welfare role of banks in the smooth-utility model, then I comment on the welfare effects of temporary monetary injections, and finally, I briefly discuss the optimal policy.

Appendix D shows that the first best is not achieved in the bankless version of the smooth-utility model if $\varepsilon^L > 0$. The logic of this result is similar to the no-bank solution in Diamond and Dybvig (1983), in which impatient households consume too little and patient households consume too much in comparison to the first best. Thus, different from the baseline model of Section 3, there is a welfare-increasing role for banks.

The effect of temporary injections on welfare depends on two forces. On the one hand, a temporary injection reduces deposits; thus, it brings the equilibrium closer to that of the bankless economy (in which the first best is not achieved, as discussed before), reducing welfare. On the other hand, a monetary injection increases money holdings M_0^h; thus, consumption of households at end of the line in a run increases because they have more money – these households withdraw $W_1^h = 0$. This second effect reduces the misallocation of consumption across impatient households at $t = 1$, which improves welfare. In the numerical examples presented here and in Appendix C.3, the second effect dominates; thus, temporary monetary injections increase welfare, although the magnitudes are small.

Finally, I comment on the optimal policy. The Friedman rule (i.e., driving the opportunity cost of holding money to zero) is optimal in my model because no agent would face a binding cash-in-advance constraint. This is because I build on the canonical framework of Lucas and Stokey (1987), in which the Friedman rule is the optimal policy as well, similar to several other monetary models (see e.g. Lucas, 2000; Lagos and Wright, 2005). Moreover, banks would play no role in my model under the Friedman rule because their only function is to offer a way to economize on the opportunity cost of holding money. Nonetheless, following Lucas (2000) and the monetary literature, my model can be extended to study the welfare cost of moving from the Friedman rule to positive inflation.[35] This exercise, however, is outside the scope of this paper because it would require calibrating a dynamic model in which banking crises happen infrequently.

7 Conclusions

A few years before the 2008 crisis, Bernanke (2002) outlined a strategy to fight deflation driven by a financial collapse. While the Federal Reserve achieved this objective in 2008, this paper argues that such strategy may generate unintended consequences: a reduction in velocity and an amplification of the flight to liquidity. The results are derived using a simple model, with the objective of clarifying the channel responsible for these unintended consequences. Two quantitative examples demonstrate the relevance of the model to studying the Great Depression and the Great Recession and show that the unintended consequences of temporary monetary injections can be large.

Some extensions could be explored in future work. If there is a positive feedback between asset prices and the condition of the banking sector, a monetary injection that increases asset prices might reduce the probability of runs. This force will contribute to reducing the flight to liquidity, offsetting my results. In contrast, if the reduction in deposits that I have identified triggers a credit crunch by reducing the resources intermediated by the banking system, macroeconomic conditions might worsen and the probability of runs might increase. This effect is likely to strengthen my results. Another possible extension relates to the tools that the central bank should use to fight deflation triggered by runs. In my model, temporary monetary injections implemented with asset purchases reduce the resources intermediated by banks. In contrast, in a framework in which central bank's loans to banks are explicitly modeled, an appropriate mix of asset purchases and loans to banks could offset deflation without shrinking the amount of resources intermediated by banks. Regardless of the extension, the channel that I have identified is likely to operate even in richer models. Finally, the model that I have developed can be extended to perform welfare analyses of monetary policies in the event of bank runs – a topic that has not received much attention in the literature.

[35] The literature typically reports the welfare cost of moving to 10% inflation.

References

Al-Najjar, Nabil I. 2004. “Aggregation and the law of large numbers in large economies.” *Games and Economic Behavior* 47 (1):1–35.

Allen, Franklin, Elena Carletti, and Douglas Gale. 2013. “Money, financial stability and efficiency.” *Journal of Economic Theory* .

Allen, Franklin and Douglas Gale. 1998. “Optimal financial crises.” *The Journal of Finance* 53 (4):1245–1284.

Alvarez, Fernando, Andrew Atkeson, and Chris Edmond. 2009. “Sluggish responses of prices and inflation to monetary shocks in an inventory model of money demand.” *The Quarterly Journal of Economics* :911–967.

Angeloni, Ignazio and Ester Faia. 2013. “Capital regulation and monetary policy with fragile banks.” *Journal of Monetary Economics* .

Antinolfi, Gaetano, Elisabeth Huybens, and Todd Keister. 2001. “Monetary stability and liquidity crises: The role of the lender of last resort.” *Journal of Economic Theory* 99 (1):187–219.

Begenau, Juliane, Saki Bigio, and Jeremy Majerovitz. 2016. “Lessons from the financial flows of the great recession.” *Working Paper* .

Bernanke, Ben S. 2002. “Deflation: making sure "it" doesn’t happen here. Remarks before the National Economists Club, Washington, D.C.” .

Brunnermeier, Markus K. and Yuliy Sannikov. 2011. “The I Theory of Money.” *Working Paper* .

Carapella, Francesca. 2012. “Banking panics and deflation in dynamic general equilibrium.” *Working Paper* .

Christiano, Lawrence J and Martin Eichenbaum. 1992. “Liquidity Effects and the Monetary Transmission Mechanism.” *The American Economic Review* 82 (2):346–353.

Cooper, Russell and Dean Corbae. 2002. “Financial collapse: A lesson from the great depression.” *Journal of Economic Theory* 107 (2):159–190.

Diamond, Douglas W. and Philip H. Dybvig. 1983. “Bank Runs, Deposit Insurance, and Liquidity.” *The Journal of Political Economy* 91 (3):401–419.

Diamond, Douglas W. and Raghuram G. Rajan. 2006. “Money in a theory of banking.” *The American Economic Review* 96 (1):30–53.

Duygan-Bump, Burcu, Patrick Parkinson, Eric Rosengren, Gustavo A Suarez, and Paul Willen. 2013. “How Effective Were the Federal Reserve Emergency Liquidity Facilities? Evidence from the Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility." *The Journal of Finance* 68 (2):715–737.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos. 2017. "Deposit competition and financial fragility: Evidence from the us banking sector." *The American Economic Review* 107 (1):169–216.

Friedman, Milton and Anna Jacobson Schwartz. 1963. *A monetary history of the United States, 1867-1960*. Princeton University Press.

———. 1970. "Monetary statistics of the United States: Estimates, sources, methods." *NBER Books* .

Gertler, Mark and Nobuhiro Kiyotaki. 2015. "Banking, Liquidity and Bank Runs in an Infinite Horizon Economy." *American Economic Review* 105 (7):2011–43.

Goldstein, Itay and Ady Pauzner. 2005. "Demand–deposit contracts and the probability of bank runs." *The Journal of Finance* 60 (3):1293–1327.

Jacklin, Charles J. 1987. "Demand deposits, trading restrictions, and risk sharing." *Contractual arrangements for intertemporal trade* 1.

Jacklin, Charles J. and Sudipto Bhattacharya. 1988. "Distinguishing panics and information-based bank runs: Welfare and policy implications." *The Journal of Political Economy* :568–592.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen. 2012. "The aggregate demand for Treasury debt." *Journal of Political Economy* 120 (2):233–267.

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

Lucas, Robert E. 2000. "Inflation and welfare." *Econometrica* 68 (2):247–274.

Lucas, Robert E. and Nancy L. Stokey. 1987. "Money and Interest in a Cash-in-Advance Economy." *Econometrica* :491–513.

Magill, Michael and Martine Quinzii. 1992. "Real effects of money in general equilibrium." *Journal of Mathematical Economics* 21 (4):301–342.

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

Nagel, Stefan. 2016. "The liquidity premium of near-money assets." *The Quarterly Journal of Economics* :1927–1971.

Robatto, Roberto. 2015. "Financial Crises and Systemic Bank Runs in a Dynamic Model of Banking." *Unicredit & Universities Working Paper Series No. 62* .

Rochet, Jean-Charles and Xavier Vives. 2004. “Coordination failures and the lender of last resort: was Bagehot right after all?” *Journal of the European Economic Association* 2 (6):1116–1147.

Rocheteau, Guillaume, Pierre-Olivier Weill, and Tsz-Nga Wong. 2015. “A tractable model of monetary exchange with ex-post heterogeneity.” *Working Paper* .

Schmidt, Lawrence, Allan Timmermann, and Russ Wermers. 2016. “Runs on money market mutual funds.” *The American Economic Review* 106 (9):2625–2657.

Wallace, Neil. 1988. “Another attempt to explain an illiquid banking system: The Diamond and Dybvig model with sequential service taken seriously.” *Federal Reserve Bank of Minneapolis Quarterly Review* 12 (4):3–16.

Appendix

A Restrictions on withdrawals at $t = 1$: discussion

In this Appendix, I comment on the three assumptions governing withdrawals at $t = 1$ introduced in Section 4.1. For each of them, I explain the role played in the rest of the paper and provide a justification related to the underlying structure of the model.

The sequential service constraint in Item (a) is crucial to obtain a flight to liquidity at $t = 0$ and, thus, to replicate this key stylized fact of the Great Depression and the Great Recession. To see this, note that a bank hit by ψ^L at $t = 1$ loses all its capital K_0^b and thus is subject to a run. The sequential service constraint implies that only some depositors are able to withdraw in the event of a run. As a result, households fly to money at $t = 0$ in order to self-insure against the risk of being at the end of the line during a run. In contrast, without the sequential service constraint, the money held by the bank would be equally split among all depositors; that is, the liquidity value of deposits (i.e., the ability to transform deposits into money at $t = 1$) would be higher. The higher liquidity value of deposits would increase their demand at $t = 0$; that is, the flight away from deposits and toward money at $t = 0$ would be more muted. Solving the model without the sequential service constraint, I obtain that households would not fly to money at all at $t = 0$; therefore, the model would not replicate the flight to liquidity observed in the data. The sequential service constraint can be imposed as a physical constraint in the environment, as in Wallace (1988), rather than as a restriction on contracts.

The no-haircut restriction in Item (b) is also required to obtain a flight to liquidity. The optimal contract would require banks hit by ψ^L to impose a haircut on deposits in order to elicit a truthful revelation of households' preference shocks, so that the bank would be able to pay some money to all impatient households. Similar to Item (a), the liquidity value of deposits would be higher, and households would not fly to money at $t = 0$. Different from Item (a), the no-haircut assumption is a restriction on contracts, although it can be motivated by the exogenous market incompleteness that precludes insurance against the idiosyncratic shocks to capital. If withdrawals at $t = 1$ could be made contingent on the realization of the idiosyncratic shocks to capital, banks would be able to offer contracts that violate the assumption of incomplete markets.

Item (c), which restricts withdrawals W_1^h to the set $\{0, D_0^h\}$, simplifies the analysis because it allows me to solve the equilibrium by conjecturing that impatient households withdraw $W_1^h = D_0^h$. The conjecture is then verified if the following incentive compatibility constraint holds:

$$\max_{C_1^h \leq \frac{M_0^h + D_0^h}{P_1}} \left\{ u\left(C_1^h\right) + \beta \frac{\overbrace{M_0^h + D_0^h - P_1 C_1^h}^{\text{unspent money}}}{\overline{P}_2} \right\} \geq$$

$$\max_{C_1^h \leq M_0^h} \left\{ u\left(C_1^h\right) + \beta \frac{\overbrace{M_0^h - P_1^h C_1^h}^{\text{unspent money}} + D_0^h \left[1 + r_2^b\left(\psi^H\right)\right]}{\overline{P}_2} \right\}. \quad (30)$$

That is, the incentive compatibility constraint in Equation (30) implies that withdrawing $W_1^h = D_0^h$ at $t = 1$, consuming $C_1^h \leq \left(M_0^h + D_0^h\right)/P_1$, and carrying any unspent money to $t = 2$ (left-hand side) gives more utility to an impatient household than withdrawing $W_1^h = 0$, consuming $C_1^h \leq M_0^h/P_1$, carrying any unspent money to $t = 2$, and getting back deposits D_0^h plus the return $r_2^b\left(\psi^H\right) > 0$ at $t = 2$ (right-hand side). The restriction $W_1^h \in \{0, D_0^h\}$ in Item (c) can be justified by the costs of contacting the bank multiple times. That is, if the household has access to the bank only once, either at $t = 1$ or at $t = 2$, restricting the choice of withdrawals to the set $\{0, D_0^h\}$ is actually optimal.[36]

[36] In addition, if I allow $W_1^h \in \left[0, D_0^h\right]$, impatient households that hold deposits at a solvent bank might choose to withdraw only a fraction of their deposits, $W_1^h < D_0^h$. This is because a solvent bank is hit by the shock to capital $\psi^b = \psi^H$, and thus, its return on deposits $r_2^b\left(\psi^H\right)$ is large (see Equations (4) and (16)). As

B Proofs

B.1 Proof of Proposition 5.1

Proof of Proposition 5.1. I start with the analysis of the household problem, taking as given equilibrium prices and the contract offered by banks. I conjecture (and later verify) that households truthfully report their type and, thus, withdraw $W_1^h = D_0^h$ if and only if they are impatient. I also consider the case in which no money is unspent at $t = 1$ (otherwise the household could do better by investing more in capital at $t = 0$, because of the positive opportunity cost of money represented by the return on capital). Thus, the cash-in-advance constraint (15) holds with equality, implying $C_1^h = \left(M_0^h + D_0^h\right)/P_1$. The problem of households is thus

$$\max_{M_0^h, K_0^h, D_0^h} \kappa \left\{ u\left(\frac{M_0^h + D_0^h}{P_1}\right) + \beta \frac{Q_0 K_0^h \left(1 + r_2^K\right)}{\overline{P}_2} \right\} + (1-\kappa)\,\beta \frac{M_0^h + D_0^h\left(1 + r_2^b\right) + Q_0 K_0^h \left(1 + r_2^K\right)}{\overline{P}_2} \tag{31}$$

subject to the non-negativity constraints $M_0^h \geq 0$, $D_0^h \geq 0$, $K_0^h \geq 0$ and to the budget constraint

$$M_0^h + D_0^h + Q_0 K_0^h \leq \overline{M} + Q_0 \overline{K}.$$

The first-order conditions imply that the non-negativity constraint on money is binding and, thus, $M_0^h = 0$. Moreover, the first-order conditions for money and capital imply

$$\kappa u'\left(\frac{D_0^h}{P_1}\right)\frac{1}{P_1} + (1-\kappa)\,\beta \frac{1 + r_2^b}{\overline{P}_2} = \beta \frac{1 + r_2^K}{\overline{P}_2}.$$

Note that $u'\left(\frac{D_0^h}{P_1}\right) = 1$ using the equilibrium values $D_0^h = D^* = \overline{M}/\kappa$ and $P_1 = P^*$ in Proposition 5.1, the functional form of $u(\cdot)$ in (2), and the restriction on $\overline{C}$ in (6). Thus:

$$\kappa \frac{1}{P_1} + (1-\kappa)\,\beta \frac{1 + r_2^b}{\overline{P}_2} = \beta \frac{1 + r_2^K}{\overline{P}_2},$$

a result, it might be convenient to leave some deposits in the bank and earn such large return. The restriction $W_1^h \in \left\{0, D_0^h\right\}$ is a simple approach that guarantees that impatient households withdraw all their deposits at $t = 1$.

which holds, given the equilibrium prices in Proposition 5.1 and the restriction on $\overline{P}_2$ in Equation (5). Thus, the allocation in Proposition 5.1 maximizes households' utility.

The conjecture that households truthfully reveal their own type can be verified as follows. First, patient households have no incentive to misreport their type, because the return from not withdrawing is positive, $r_2^b > 0$. Second, the incentive compatibility constraint in Equation (30) holds when evaluated at $M_0^h = 0$, $D_0^h = D^*$, and $P_1 = P^*$, and using the functional form of $u(\cdot)$ in (2) and the restriction on $\overline{C}$ in (6). Capital K_0^h is residually determined by the budget constraint and consumption at $t = 2$ follows from (17).

Market clearing for money holds trivially because $M_0^h = 0$ and $M_0^b = \kappa D_0^h = \overline{M}$. Market clearing for consumption goods also holds because consumption by impatient households at $t = 1$ is $C_1^h = A\overline{K}/\kappa$ and there is a mass fraction of them; thus, that total consumption is $A\overline{K}$ and equals output. The market-clearing condition for capital holds by Walras' Law.

Banks must invest at least a fraction κ of money in order to serve withdrawals by impatient households. Moreover, it is not optimal to offer contracts that specify $M_t^b > \kappa D_t^b$; a bank that does so invests less in capital, and thus, the return r_2^b would be lower in comparison to the return offered by other banks that choose $M_t^b = \kappa D_t^b$. Therefore, banks choose $M_0^b = \kappa D_0^b$. Capital K_0^b is residually determined by the budget constraint. □

B.2 Proof of Proposition 5.3

This section is organized as follows. I first provide more details on the maximization problem of households. I then restate Proposition 5.3 by specifying the equilibrium values of prices and allocations as functions of parameters and policy variables. Finally, I prove the results.

The household problem at $t = 0$ is given by

$$\max_{M_0^h, D_0^h, K_0^h} \underbrace{(1-\alpha)}_{\text{Case 1}} \left[\kappa \underbrace{u\left(\frac{M_0^h + D_0^h}{P_1}\right)}_{\text{impatient}} + (1-\kappa)\underbrace{\beta \frac{D_0^h\left(1 + r_2^b\left(\psi^H\right)\right) + M_0^h}{\overline{P}_2}}_{\text{patient}} \right] \tag{32}$$

$$+ \underbrace{\alpha \times Pr\,(\text{beginning of line})}_{\text{Case 2}} \left[\kappa \underbrace{u\left(\frac{M_0^h + D_0^h}{P_1}\right)}_{\text{impatient}} + (1-\kappa)\underbrace{\beta \frac{M_0^h + D_0^h}{\overline{P}_2}}_{\text{patient}} \right]$$

$$+\underbrace{\alpha \times Pr\left(\text{end of line}\right)}_{\text{Case 3}}\left[\kappa\,\underbrace{u\left(\frac{M_0^h}{P_1}\right)}_{\text{impatient}}+\left(1-\kappa\right)\beta\underbrace{\frac{M_0^h}{\overline{P}_2}}_{\text{patient}}\right]+\beta\frac{Q_0K_0^h\mathbb{E}\left\{1+r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2}$$

subject to the budget constraint in Equation (13). At $t=1$, a household h faces three cases. With probability $1-\alpha$, the bank of household h is hit by $\psi^H>0$ and is thus solvent and not subject to a run; therefore, an impatient household can withdraw $W_1^h=D_0^h$ (Case 1).[37] With probability α, the bank of household h is hit by the negative shock to capital, ψ^L, and is thus subject to a run. The household can be either at the beginning of the line and thus able to withdraw $W_1^h=D_0^h$ (Case 2) or at the end of the line and thus unable to withdraw, $W_1^h=0$ (Case 3). I conjecture that the household makes the following choices and then verify the conjecture in the proof of the proposition:

a. In Case 1, if h is impatient (i.e., with probability κ), it withdraws $W_1^h=D_0^h$ and consumes $C_1^h=\left(M_0^h+D_0^h\right)/P_1$; if it is patient (i.e., with probability $1-\kappa$), it withdraws and consumes zero, carrying M_0^h to $t=2$.

b. In Case 2, the household withdraws $W_1^h=D_0^h$ no matter what its preference shock is; it then consumes $C_1^h=\left(M_0^h+D_0^h\right)/P_1$ if it is impatient (i.e., with probability κ) and zero if patient (i.e., with probability $1-\kappa$), carrying $M_0^h+D_0^h$ to $t=2$.

c. In Case 3, the household cannot withdraw; if h is impatient (i.e., with probability κ), it faces a very tight cash-in-advance constraint and can consume only $C_1^h=M_0^h/P_1$; if h is patient (i.e., with probability $1-\kappa$), it consumes zero and carries M_0^h to $t=2$. Regardless of whether the household is patient or impatient, it will lose all of its deposits because $r_2^b\left(\psi^L\right)=-1$.

Because of the piecewise-linear utility function, the first-order conditions of (32) are linear functions of prices and do not depend on households' choices.[38] Thus, such first-order conditions combined with the market-clearing conditions in Section 4.3 form a linear system of equations with a unique solution.

The probability that a household is at the beginning or at the end of a line in the event of a run depends on the fraction of deposits that banks invest in money at $t=0$; the higher is the fraction of money at $t=0$, the higher is the fraction of households that can be served in the event of a run.

[37] In the proof of Proposition 5.3, I verify that an impatient household holding deposits at a solvent bank prefers to withdraw $W_1^h=D_0^h$ instead of $W_1^h=0$.

[38] More precisely, the first-order conditions equated to zero are linear functions of $1/P_1$ and $r_2^b\left(\psi^H\right)$, given $\overline{P}_2$.

From the analysis of the household problem in (32), it follows that households' consumption is not equalized in the three cases. That is, households that face Cases 1 and 2 at $t = 1$ consume more than households that face Case 3. The different consumption among impatient households gives rise to a welfare loss because marginal utilities are not equalized across such households. In addition, consumption in Case 3 is crucial to give rise to the flight to liquidity at $t = 0$. In equilibrium, consumption in Case 3 is less than $\overline{C}$. Since $u(\cdot)$ is characterized by global risk aversion, households fly to money at $t = 0$ to partially self-insure against the risk of facing Case 3 and thus consuming $C_1^h < \overline{C}$. If $u(\cdot)$ were globally linear, households would be risk-neutral with respect to time-1 consumption, and thus, consumption risk at $t = 1$ would be irrelevant, implying no flight to liquidity at $t = 0$.

I now restate Proposition 5.3 by specifying the equilibrium values of prices and allocations as functions of parameters and policy variables, and then I prove the results.

Proposition. *Fix the money supply $M_0^S = M_2^S = \overline{M}$. If the parameters satisfy:*[39]

$$0 < (1-\alpha)(1-\beta\kappa) - \alpha\kappa(\theta-1) < (1-\alpha)^2 \beta (1-\kappa), \tag{33}$$

$$\theta > 1 + \frac{\alpha(1-\alpha)}{1 - 3\alpha(1-\alpha) + \alpha\kappa - \alpha^2\kappa(1+\kappa) - \alpha^3(1-\kappa)}, \tag{34}$$

there exists an equilibrium with bank runs characterized by the following prices and allocation:

- *Prices are*

$$Q_0 = \frac{\overline{M}}{\overline{K}} \frac{[1-\alpha-\alpha\kappa(\theta-1)]\left[(1-\beta)(1-\alpha-\alpha\kappa(\theta-1)) + \beta^2(1-\alpha)\right]}{(1-\beta)\beta(1-\alpha)\left[1-\alpha-\alpha^2\kappa(1-\kappa)(\theta-1)\right]} < Q^* \tag{35}$$

  (where the inequality $Q_0 < Q^*$ holds if β is sufficiently close to one) and

$$P_1 = \frac{\overline{M}\left[1-\alpha(1+\kappa(\theta-1))\right]}{A_1\overline{K}\beta(1-\alpha)} < P^*. \tag{36}$$

- *$t = 0$: deposits and money holdings are*

$$D_0^h = D_0^b = \overline{M}\left[\frac{(1-\alpha)(1-\beta\kappa) - \alpha\kappa(\theta-1)}{(1-\alpha)^2\beta(1-\kappa)\kappa}\right],$$

[39] The restriction in (34) requires θ to be large enough that households fly to liquidity. That is, if (34) holds, the marginal utility of households that consume less than $\overline{C}$ – in equilibrium, those at the end of the line in a run – is sufficiently large that households fly to liquidity.

$$
\begin{aligned}
M_0^b &= \overline{M}\left[\frac{(1-\alpha)(1-\beta\kappa)-\alpha\kappa(\theta-1)}{(1-\alpha)^2\beta(1-\kappa)}\right], \\
M_0^h &= \overline{M}\left[1-\frac{(1-\alpha)(1-\beta\kappa)-\alpha\kappa(\theta-1)}{(1-\alpha)^2\beta(1-\kappa)}\right];
\end{aligned}
$$

holdings of capital, K_0^h and K_0^b, are residually determined by the budget constraints of households and banks, respectively:

$$
K_0^b=\frac{\overline{K}(1-\beta)[1-\alpha(1+\alpha(\theta-1)(1-\kappa)\kappa)][\alpha(\kappa(\beta-\theta+1)-1)+1-\beta\kappa]}{(1-\alpha)\kappa[1-\alpha-\alpha(\theta-1)\kappa][(1-\alpha)(1-(1-\beta)\beta)-\alpha(1-\beta)(\theta-1)\kappa]},
$$

$$
K_0^h=\overline{K}\left(1-\frac{(1-\beta)[1-\alpha(1+\alpha(\theta-1)(1-\kappa)\kappa)][\alpha(\kappa(\beta-\theta+1)-1)+1-\beta\kappa]}{(1-\alpha)\kappa[1-\alpha-\alpha(\theta-1)\kappa][(1-\alpha)(1-(1-\beta)\beta)-\alpha(1-\beta)(\theta-1)\kappa]}\right);
$$

- $t=1$*: banks hit by ψ^L are subject to runs, whereas banks hit by ψ^H are not subject to runs; that is, for depositors of banks hit by ψ^L, withdrawals are*

$$
W_1^h=\begin{cases} D_0^h & \textit{for a fraction } \kappa \textit{ of depositors} \\ 0 & \textit{for a fraction } 1-\kappa \textit{ of depositors,} \end{cases}
$$

whereas for depositors of banks hit by ψ^H,

$$
W_1^h=\begin{cases} D_0^h & \textit{if } h \textit{ is impatient} \\ 0 & \textit{if } h \textit{ is patient}; \end{cases}
$$

consumption is:

$$
C_1^h=\begin{cases} \left(M_0^h+W_1^h\right)/P_1 & \textit{if } h \textit{ is impatient} \\ 0 & \textit{if } h \textit{ is patient.} \end{cases}
$$

- $t=2$*: the return on deposits not withdrawn is*

$$
r_2^b\left(\psi^b\right)=\begin{cases} \frac{\alpha}{1-\alpha}\left[1+\frac{\kappa(\theta-1)[1-\alpha(1-\kappa)]}{1-\alpha(1+\kappa(\theta-1))}\right] & \textit{if } \psi^b=\psi^H \\ -1 & \textit{if } \psi^b=\psi^L. \end{cases} \tag{37}
$$

Proof. Consider the households' problem in (32). As discussed in Section 5.2, the marginal utility of consumption in Cases 1 and 2 is one, whereas the marginal utility is $\theta>1$ in Case

3. I conjecture (and later verify) that Pr (beginning of line) $= \kappa$ and Pr (end of line) $= 1-\kappa$. Using this conjecture, the household problem implies three households' first-order conditions (with respect to M_0^h, K_0^h, and D_0^h) that are independent of consumption allocation, as discussed in Section 5.2, and linear in $r_2^b\left(\psi^H\right)$ and $1/P_1$. Thus, these equations allow me to solve for $r_2^b\left(\psi^H\right)$ in (37) and

$$P_1 = \frac{M_2^S\left[1-\alpha\left(1+\kappa\left(\theta-1\right)\right)\right]}{A_1\overline{K}\beta\left(1-\alpha\right)} \tag{38}$$

in addition to the Lagrange multiplier of the budget constraint. Evaluating (38) at $M_2^S = \overline{M}$, I obtain (36); the inequality in (36) follows from (33). The $r_2^b\left(\psi^H\right)$ and P_1 are both positive because of the assumption in (33) and $\beta > 0$, $\kappa > 0$, $0 < \alpha < 1$. Moreover, (36), (37), and Equations (4) and (16) evaluated at $\psi^b = \psi^H$ allow me to solve for Q_0 in (35). The result $Q_0 < Q^*$ holds because, using (33):

$$\begin{aligned} Q_0 &= \frac{\overline{M}}{\overline{K}}\frac{\left[1-\alpha-\alpha\kappa\left(\theta-1\right)\right]\left[\left(1-\beta\right)\left(1-\alpha-\alpha\kappa\left(\theta-1\right)\right)+\beta^2\left(1-\alpha\right)\right]}{\left(1-\beta\right)\beta\left(1-\alpha\right)\left[1-\alpha-\alpha^2\kappa\left(1-\kappa\right)\left(\theta-1\right)\right]} \\ &< \frac{\overline{M}}{\overline{K}}\frac{\left(1-\beta\right)\left(1-\alpha-\alpha\kappa\left(\theta-1\right)\right)/\left(\beta(1-\alpha)\right)+\beta}{\left(1-\beta\right)} \end{aligned}$$

and because β can be taken to be arbitrarily close to one, so that the expression in the second line is arbitrarily close to $\left(\overline{M}/\overline{K}\right)\beta/\left(1-\beta\right)$, that is, arbitrarily close to Q^*.

Next, I solve for M_0^h, D_0^h, and M_0^b. To do so, I use the market-clearing conditions (20) and (21) and the banks' choice of money holdings $M_0^b = \kappa D_0^b$ (I will verify the optimality of this choice later). The market-clearing condition for money holdings and the fact that all banks are alike and all households are alike imply $M_0^h + M_0^b = M_0^S$. This equation is also linear in the endogenous variables. The market-clearing condition for consumption goods, Equation (21), can be used to pin down the price level, similar to the bankless economy (Section 3). Multiplying both sides of Equation (21) by P_1, I obtain $\int P_1 C_1^h dh = P_1 A_1 \overline{K}$, where $P_1 C_1^h$ is the consumption expenditure of household h. To compute the consumption expenditure of households, I use the three cases analyzed in Section 5.2 and the law of large numbers. A fraction $1-\alpha$ of households have deposits at banks not subject to runs and, thus, spend $M_0^h + D_0^h$; a fraction $\alpha\kappa$ have deposits at banks subject to runs but are at the beginning of the line; therefore, they spend $M_0^h + D_0^h$ as well (recall that I am conjecturing Pr (beginning of line) $= \kappa$); the remaining fraction $\alpha\left(1-\kappa\right)$ are at the end of the line and, thus, spend only M_0^h. Therefore, the market-clearing condition become:

$$P_1 A_1 \overline{K} = (1 - \alpha + \alpha\kappa)\left(M_0^h + D_0^h\right) + \alpha(1-\kappa) M_0^h. \tag{39}$$

Banks' choice of money holdings $M_0^b = \kappa D_0^b$ implies $M_0^b = \kappa D_0^h$, using the deposit market clearing condition in (20).

To sum up, I have three linear equations in M_0^h, D_0^h, and M_0^b: $M_0^h + M_0^b = M_0^S$, (39), and $M_0^b = \kappa D_0^h$. These equations imply

$$D_0^h = \frac{M_2^S\left[1 - \alpha - \alpha\kappa(\theta - 1)\right] - M_0^S(1-\alpha)\beta\kappa}{(1-\alpha)^2\beta(1-\kappa)\kappa} \tag{40}$$

$$M_0^b = \frac{M_2^S\left[1 - \alpha - \alpha\kappa(\theta - 1)\right] - M_0^S(1-\alpha)\beta\kappa}{(1-\alpha)^2\beta(1-\kappa)}$$

$$M_0^h = \frac{M_0^S(1-\alpha)\beta\left[1 - \alpha(1-\kappa)\right] - M_2^S\left[1 - \alpha - \alpha\kappa(\theta - 1)\right]}{(1-\alpha)^2\beta(1-\kappa)\kappa} \tag{41}$$

which simplify to the results stated in the proposition when evaluated at $M_0^S = \overline{M}$ and $M_2^S = \overline{M}$. Given the restriction on parameters in (33), D_0^h, M_0^b, and M_0^h are positive.

To prove the results about consumption C_1^h of impatient households at $t = 1$, I use the households' objective function, (32). Households that face Cases 1 and 2 at $t = 1$ consume more than households that face Case 3. Moreover, since all impatient households were consuming $\overline{C}$ in the economy with no runs, it must be the case that households facing Cases 1 and 2 consume more than $\overline{C}$ (and thus their marginal utility is one), whereas households facing Case 3 consume less than $\overline{C}$ (and thus their marginal utility is $\theta > 1$).

The quantity of capital held by banks and households, K_0^b and K_0^h, is given residually by the respective budget constraints. The market-clearing condition for capital holds by Walras' Law.

To conclude the analysis of the households' problem, I must verify two guesses. First, to verify that $Pr\,(\text{beginning of line}) = \kappa$, note that all households and all banks are identical; therefore, a bank subject to a run serves a fraction of depositors equal to the ratio of money M_0^b to deposits D_0^b. Since $M_0^b = \kappa D_0^b$, the result follows. Second, the households' problem in (32) is formulated under the guess that households truthfully reveal their own types in Case 1. For impatient households, I verify the conjecture by checking that the incentive constraint in Equation (30) holds when evaluated at the equilibrium values of the endogenous variables, using the restriction on parameters in (33) and (34). For patient households, the conjecture is verified because $r_2^b\left(\psi^H\right)>0$; thus, the return from waiting until $t=2$ is higher than the return on withdrawing money and carrying the money to $t=2$ (which is zero).

To verify the optimality of $M_0^b=\kappa D_0^b$, note first that M_0^b must be greater than or equal to κD_0^b in order to serve withdrawals by impatient households if the bank is not subject to a run. Thus, I need to verify that $M_0^b>\kappa D_0^b$ is not optimal by showing that $M_0^b>\kappa D_0^b$ reduces households' utility, so that such contract is not offered in equilibrium. In order to offer the best contract among the class of contracts that I allow, the bank chooses the composition of money and capital, M_0^b and K_0^b, in order to maximize the objective function of households:

$$\max_{M_0^b,K_0^b}(1-\alpha)\left[\kappa u\left(\frac{M_0^h+D_0^h}{P_1}\right)+(1-\kappa)\beta\frac{D_0^h\left(1+\hat{r}_2^b\left(\psi^H\right)\right)+M_0^h}{\overline{P}_2}\right] \tag{42}$$
$$+\alpha\frac{M_0^b}{D_0^b}\left[\kappa u\left(\frac{M_0^h+D_0^h}{P_1}\right)+(1-\kappa)\beta\frac{M_0^h+D_0^h}{\overline{P}_2}\right]$$
$$+\alpha\left(1-\frac{M_0^b}{D_0^b}\right)\left[\kappa u\left(\frac{M_0^h}{P_1}\right)+(1-\kappa)\beta\frac{M_0^h}{\overline{P}_2}\right]+\beta\frac{Q_0K_0^h\mathbb{E}\left\{1+r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2}$$

subject to the budget constraint of the bank, (12), and where

$$1+\hat{r}_2^b\left(\psi^H\right)=\frac{\left(D_0^b-M_0^b\right)\left(1+r_2^K\left(\psi^b\right)\right)+M_0^b-\kappa D_0^b}{(1-\kappa)D_0^b}. \tag{43}$$

Note that M_0^b/D_0^b in (42) is the probability that an household is at the beginning of the line in the event of a run, as discussed before. Moreover, (43) collapses to (16) if $M_0^b=\kappa D_0^b$; if instead $M_0^b>\kappa D_0^b$, fewer resources are invested in capital at $t=0$, but $M_0^b-\kappa D_0^b$ is not used to pay withdrawals at $t=1$ and, thus, is left to repay deposits not withdrawn at $t=2$.

Plugging (12) and (43) into (42), I obtain an unconstrained problem in M_0^b. The first-order condition, evaluated at $M_0^b=\kappa D_0^b$ and at the equilibrium values of the other endogenous variables, is

$$-\frac{A_1\overline{K}\alpha(\theta-1)}{(1-\alpha)}\left[\frac{(1-\alpha)(1-\beta\kappa)-\alpha\kappa(\theta-1)+(1-\alpha)^2\beta\kappa}{1-\alpha-\alpha\kappa(\theta-1)}\right]<0$$

where the inequality follows from (33); that is, the first-order condition evaluated at $M_0^b=\kappa D_0^b$ is negative, so that increasing M_0^b above κD_0^b reduces households' utility.

Finally, I show that offering a run-proof contract is not optimal either. To have a run-proof contract, the bank must invest 100% of its deposits into money, $M_0^b = D_0^b$ (if instead the bank invest some resources in capital, the bank will be insolvent in the event of a bad shock $\psi^L = -1$ at $t = 1$ and thus subject to a run) and pays no return on deposits. Thus, a run-proof contract is equivalent to an household that decides to hold no deposits and $M_0^h > 0$, $K_0^h > 0$. To show that this deviation is not welfare-increasing for household, consider the problem:

$$\max_{M_0^h, K_0^h} \kappa \left\{ u\left(\frac{M_0^h}{P_1}\right) + \beta \frac{Q_0 K_0^h \, \mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2} \right\} + (1-\kappa)\,\beta \frac{M_0^h + Q_0 K_0^h \, \mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2}$$

subject to the budget constraint in (13) evaluated at $D_0^h = 0$. This problem is similar to (7), but prices are different. The first-order conditions imply the same expression as in (10), but evaluated possibly with inequality. Using the equilibrium prices of the economy with runs, the left-hand side of (10) is greater than the right-hand side if $u'\left(M_0^h/P_1\right) = 1$, whereas the left-hand side is less than the right-hand side if $u'\left(M_0^h/P_1\right) = \theta$. As a result, the optimal choice of money is the one that allows the household to consume $C_1^h = \overline{C}$ at $t = 1$, that is, $M_0^h = P_1\overline{C}$. I then compute household's utility under the deviation $D_0^h = 0$ and $M_0^h = P_1\overline{C}$ and compare it with household's utility evaluated at the choices of the economy with runs. The difference between the two utilities, both evaluated at the same equilibrium prices, is $-A_1\overline{K}\alpha\,(\theta-1)\,(1-\kappa) < 0$; that is, the deviation leads to a loss in utility and thus it is not profitable. □

B.3 Proof of the results in Section 5.3

Proof of Proposition 5.4. The results for P_1, Q_0, D_0^h, and M_0^h follow by differentiating Equations (38), (35), (40), and (41) with respect to M_0^S. In particular, the signs of the elasticities of D_0^h and M_0^h follow from

$$\frac{dD_0^h}{dM_0^S} = -\frac{1}{(1-\alpha)(1-\kappa)} < 0$$

$$\frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} = \frac{M_0^S(1-\alpha)\,\beta\,(1-\alpha(1-\kappa))}{M_0^S(1-\alpha)\,\beta\,(1-\alpha(1-\kappa)) - M_2^S(1-\alpha-\alpha\kappa(\theta-1))} > 1$$

where the last inequality uses the restriction on parameters in (33). The result for money velocity follows using the definition of v in Equation (24) and rearranging. □

Proof of Proposition 5.5. The results follow by setting $M_0^S = M_2^S = \widehat{M}$ in Equations (5), (24), (38), (35), (40), and (41), and differentiating with respect to $\widehat{M}$. □

Proof of Proposition 5.6. The results follow by differentiating Equations (5), (24), (38), (35), (40), and (41) with respect to M_2^S and using the restriction on parameters in (33). □

Proof of Proposition 5.7. If $\alpha = 0$ and $M_2^S = \overline{M}$, the equilibrium is characterized by

$$M_0^h = 0, \quad D_0^h = \frac{M_0^S}{\kappa}, \quad P_1 = \frac{M_0^S}{A_1\overline{K}}, \quad r_2^K = \frac{\overline{M}}{M_0^S}\beta - 1 \tag{44}$$

as long as $r_2^K > 0$ (i.e., as long as $M_0^S < \overline{M}/\beta$). That is, a positive r_2^K implies a positive opportunity cost of holding money so that households' optimal level of money holdings is $M_0^h = 0$; the other results in (44) can be proven in a way similar to Proposition 5.1 (indeed, setting $M_0^S = \overline{M}$, the values in (44) become the same as in Proposition 5.1).

The results of the proposition follow by differentiating (44) and velocity v, defined using Equation (24), with respect to M_0^S. □

B.4 Proof of Proposition 6.1

I start by analyzing the problem of households, which is similar to (31). However, patient households may decide to consume since $\varepsilon^L \geq 0$. I conjecture that patient and impatient households consume, respectively,

$$C_1\left(\varepsilon^L\right) = \frac{M_0^h}{P_1}, \qquad C_1\left(\varepsilon^H\right) = \frac{D_0^h + M_0^h}{P_1}. \tag{45}$$

This conjecture is then verified because it is not optimal to have any unspent money at $t = 1$ because of the positive opportunity cost of holding money represented by the return on capital. As a result, the household problem is

$$\begin{aligned}\max_{M_0^h, D_0^h, K_0^h} \kappa \left\{\varepsilon^H u\left(\frac{M_0^h + D_0^h}{P_1}\right) + \beta\frac{Q_0 K_0^h\left(1 + r_2^K\right)}{P_2}\right\}& \\ + (1-\kappa)\left\{\varepsilon^L u\left(\frac{M_0^h}{P_1}\right) + \beta\frac{Q_0 K_0^h\left(1 + r_2^K\right) + D_0^h\left(1 + r_2^b\right)}{P_2}\right\}&\end{aligned} \tag{46}$$

subject to the budget constraint, (12). The first-order conditions imply, using the functional form $u(C) = \overline{C} \log C$,

$$\kappa \frac{\varepsilon^H}{M_0^h + D_0^h} + (1-\kappa) \frac{\varepsilon^L}{M_0^h} = \left(\frac{A\overline{P}_2 + A_1 P_1}{Q_0} \right) \beta \frac{1}{\overline{P}_2} \tag{47}$$

$$\kappa \frac{\varepsilon^H}{M_0^h + D_0^h} + (1-\kappa) \beta \frac{\left(1 + r_2^b\right)}{\overline{P}_2} = \left(\frac{A\overline{P}_2 + A_1 P_1}{Q_0} \right) \beta \frac{1}{\overline{P}_2}. \tag{48}$$

The market-clearing condition for money, together with banks' choice $M_0^b = \kappa D_0^b$ and the market-clearing condition for deposits $D_0^b = D_0^h$, implies

$$M_0^S = M_0^h + \kappa D_0^h. \tag{49}$$

The market-clearing condition for consumption at $t = 1$, multiplied by P_1, implies

$$\int P_1 C_1^h dh = P_1 A_1 \overline{K}, \tag{50}$$

where $\int P_1 C_1^h dh = P_1 \left[\kappa C_1 \left(\varepsilon^H\right) + (1-\kappa) C_1 \left(\varepsilon^L\right)\right]$.

Summing up, I have a system of ten equations (Equations (4), (16) evaluated at $\psi^H = 1$, the two guesses in (45), (47)-(50), $M_0^b = \kappa D_0^b$, and $D_0^b = D_0^h$) in ten endogenous variables (D_0^h, M_0^h, Q_0, P_1, r_2^K, $C_1\left(\varepsilon^H\right)$, $C_1\left(\varepsilon^L\right)$, D_0^b, M_0^h, r_2^b). The solution, evaluated at $M_0^S = M_2^S = \overline{M}$, is given by the results stated in the proposition. The choice $M_0^b = \kappa D_0^b$ of banks can be shown to be optimal with the same approach discussed for Proposition 5.1.

C The run on MMMFs: discussion and robustness

This appendix discusses further the comparison between the model and the run on prime institutional MMMFs in September 2008.

C.1 Many states of the world at $t = 1$ and runs in the model versus the data

As noted in Section 5.2, the model can be extended by adding two aggregate states at $t = 1$: a state in which there are idiosyncratic shocks to capital, and thus, runs on banks hit by ψ^L, and a state in which the idiosyncratic shocks to capital are zero for all agents. In the second state, no bank would be subject to runs. The model would still produce a flight to liquidity at $t = 0$ (replicating the redemptions in the month after the collapse of Lehman Brothers) and, in the second aggregate state, no runs at $t = 1$ (replicating the fact that no fund was subject to a full run thereafter).

C.2 Robustness: Timing of AMLF announcement and Treasury Guarantee Program

I present a robustness analysis of the quantitative exercise related to the run on MMMFs in September 2008. The objective is to address two possible concerns, related to (i) the timing of the announcement of the AMLF (i.e., the liquidity facility set up by the Federal Reserve) and (ii) the announcement of the Guarantee Program by the Treasury to provide insurance to MMMFs.

The AMLF and Treasury guarantee announcements were made on Friday, September 19, four days after the collapse of Lehman Brothers. During these four days, approximately \$325 billion was redeemed from prime institutional MMMFs (see Figure 1, Panel B, in Schmidt, Timmermann, and Wermers, 2016). Despite the two announcements, the run continued and an additional \$100 billion was redeemed from prime institutional MMMFs from Monday, September 22, to mid-October (see Figure 1, Panel B, in Schmidt, Timmermann, and Wermers, 2016).

To deal with some possible concerns related to the timing of the AMLF and Guarantee Program announcements (which I explain next), in this appendix I restrict attention to the \$100 billion redemptions that took place after the announcement of the AMLF and of the Treasury Guarantee Program. That is, I recalibrate the model and perform a new quantitative exercise, focusing only on the events after September 19.

Policies announcements: model versus data. In the model, monetary injections are announced at time $t = 0$, before the Walrasian market opens and before depositors make any choice. However, the exact timing of the announcement is not relevant if it is anticipated. That is, even if the policy intervention is announced while some trading in the Walrasian market has already taken place, the announcement has no effect on prices or households' decisions as long as such announcement is fully anticipated. In practice, it is possible that the AMLF could have been anticipated, at least in part. The Federal Reserve had announced its commitment to provide liquidity to the financial market right after the collapse of Lehman Brothers (Federal Reserve Press Release, 09/14/2008), and spillovers to MMMFs appeared to be one of the concerns that motivated the AIG bailouts on September 16.[40]

The robustness check in this appendix deals with the possibility that the AMLF announcement was not anticipated. In this case, the $325 billion redemptions in the week of September 15-19 would have been decided under the presumption of no policy intervention by the Federal Reserve. As a result, the counterfactual policy analysis that asks, "What would have happened if the Federal Reserve had not established the AMLF?" (and thus the calibration of the model) should be conducted by focusing only on the fraction of the run that took place after the announcement of the AMLF on September 19.

The second concern addressed by the robustness check is related to the Treasury Guarantee Program. This program is akin to deposit insurance in which funds are insured in exchange for a fee. Two comments are related to this announcement. First, I claim that investors perceived a positive probability of a run even after the Guarantee Program was announced; thus, the model is still relevant for the analysis. While full deposit insurance should completely eliminate runs and flight to liquidity, the data show that the run continued after the Guarantee Program was announced; that is, an additional $100 billion was redeemed after September 19. This behavior can be rationalized if the Guarantee Program was perceived as partial insurance (i.e., insurance that would have covered losses only in part or only under some contingencies). Indeed, while commercial banks are typically required to have deposit insurance, the Guarantee Program for MMMFs left the participation decision to each fund, and funds had until October 8 to apply.[41] Moreover, the funds set aside for the Guarantee Program were limited in comparison to the size of the MMMFs industry.[42] Second, if the announcement by the Treasury was not anticipated, it might have affected the probability of a run. Therefore, restricting attention to the events after September 19 allows me to focus on a subsample in which the Treasury policy is kept constant.

[40] For the concern related to spillovers to MMMFs, see Karnitschnig, M., D. Solomon, L. Pleven, and J. E. Hilsenrath, "US to take over AIG in $85 billion bailout; central banks inject cash as credit dries up," *Wall Street Journal,* September 16, 2008.

[41] See the press release at https://www.treasury.gov/press-center/press-releases/Pages/hp1161.aspx.

[42] The Guarantee Program was funded with only $50 billion, whereas the overall size of the MMMFs industry in September 2008 was about $3.5 trillion (Duygan-Bump et al., 2013). If there is uncertainty about the number of mutual funds that can be in trouble (i.e., uncertainty about α in the model), there could be states of the world in which many funds are subject to runs, and thus, $50 billion is not sufficient to cover all the losses.

Calibration. On September 19, about \$325 billion had been redeemed from MMMFs; thus, the liabilities of MMMFs were down from \$1.3 trillion before the collapse of Lehman Brothers to $\$1,300 - \$325 = \$975$ billion on September 19. I choose the values of κ, ε^H, and $\overline{M}$ so that deposits in the economy with no runs, D^*, equate $\$975$ billion. I set $\kappa = 0.1$, a lower value compared to the calibration in Section 6.3.2. Recall that in Section 6.3.2, I set $\kappa = 0.2$ to match the amount of liquid assets held by MMMFs before the collapse of Lehman Brothers; however, the massive redemptions in the week of September 15-19 likely reduced the amount of liquidity available to MMMFs. Similar to Section 6.3.2, I set $\varepsilon^H = 1/\kappa = 10$ and $\varepsilon^L = 0$, which is the choice that produces the most conservative results of the policy analysis. I set $\overline{M} = \$97.5$ billion so that $D^* = \overline{M}/\kappa = \$975$ billion, as explained before (see Proposition 6.1 for the definition of D^*).

The M_0^S, M_2^S, and α are calibrated using the same approach as in Proposition 6.1 (i.e., to match some key facts observed during the run on MMMFs), although their values are different. I set $M_0^S = \overline{M} + \$150 = \$247.5$ billion and $M_2^S = \overline{M}$ to capture the \$150 billion temporary injections implemented by the Federal Reserve using the AMLF. I choose α to match price stability (i.e., P_1 in the economy with runs and monetary injections equals P^*). The resulting value is $\alpha = 0.0012$.

Finally, I normalize A_1 and $\overline{K}$ so that prices in the economy with no runs are $Q^* = 100$ and $P^* = 100$, and I set the discount factor to $\beta = 0.9985$. The choice of β is slightly higher compared to Proposition 6.1 to account for the shorter period under analysis.

Results. In the economy with runs, deposits are $D_0^h = \$809$ billion. Thus, redemptions from MMMFs, computed as $D_0^h - D^*$ as in Section 6.3.2, are $\$166$ billion. That is, the model overestimates the value of redemptions, which are \$100 billion in the data. One possible reason for such overestimation is that I calibrate the model by comparing the data on September 19 to the economy with no runs, whereas the data in September 19 refer to an ongoing financial crisis. The flight to money by households in the model, given by money holdings M_0^h in the economy with runs, is $\$167$ billion.

The local elasticities of the key endogenous variables with respect to an additional temporary injection are

$$\frac{dv}{dM_0^S} \times \frac{M_0^S}{v} = -0.99, \qquad \frac{dP_1}{dM_0^S} \times \frac{M_0^S}{P_1} = 0.002,$$

$$\frac{dD_0^h}{dM_0^S} \times \frac{M_0^S}{D_0^h} = -0.34, \qquad \frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} = 1.65.$$

The magnitude of these results is similar to that in (28) and (29). The price level is almost unaffected by a temporary injection, whereas the elasticity is slightly smaller for both deposits D_0^h (in absolute value) and money holdings M_0^h. Nonetheless, the main results of Section 6.3.2 are unchanged; that is, the degree of monetary non-neutrality is very high.

As in Section 6.3.2, I also present the results of the global effects of monetary injections on the flight to liquidity and the price level. That is, I ask what would have happened without the AMLF (i.e., if $M_0^S = \overline{M}$). In this case, deposits would have been $D_0^h = \$935$ billion, the flight to money by households would have been $M_0^h = \$3.95$ billion, and the price level P_1 would have dropped by 3.8% in comparison to P^*. That is, the robustness exercise confirms the global policy analysis as well. According to the model, the Federal Reserve avoided deflation in September 2008 at the expense of a substantial amplification of the flight to liquidity. The magnitude of the force that I have identified and that gives rise to the unintended consequences of monetary injections is economically important.

C.3 Alternative calibration of ε^H, ε^L

I now present the results of an alternative calibration of the preference shocks (ε^H and ε^L), and I compare the results with those derived in Section 6.3.2. In the calibration Section 6.3.2, $\varepsilon^H = 1/\kappa$ and $\varepsilon^L = 0$, whereas I now consider $\varepsilon^H < 1/\kappa$ and $\varepsilon^L > 0$. The calibration used in this appendix is less conservative because it produces larger effects of policy interventions; however, I argue that it is useful because it puts an upper bound on the magnitude of the results.

Calibration. I choose $\overline{M} = \$1,681$ billion and $\varepsilon^H = 1.62$; I use the same value of κ as in Section 6.3.2, and thus the restriction $\mathbb{E}_0\left(\varepsilon_1^h\right) = 1$ imposed in Section 6.1 implies $\varepsilon^L = 0.85$. In the economy with no runs (Proposition 6.1), these choices imply that $M_0^h = \$1,421$ billion, corresponding to the stock of M1 in the data on September 8, 2016.[43] That is, under this calibration, investors that own the shares of prime institutional MMMFs hold the entire M1 in the U.S. economy as well. This calibration likely overestimates the fraction of M1 held by investors that own the shares of prime institutional MMMFs. However, this calibration is the opposite of Section 6.3.2, in which $\varepsilon^H = 1/\kappa$, $\varepsilon^L = 0$, and investors that

[43] Source: FRED (Federal Reserve Bank of St. Louis Economic Data).

own the shares of prime institutional MMMFs do not hold any fraction of M1. Thus, the correct calibration and the results of the policy analyses are likely to be an intermediate case. The choices $\overline{M} = \$1,681$ billion and $\varepsilon^H = 1.62$ also imply that deposits in the economy with no runs are $D^* = \$1.3$ trillion (see Proposition 6.1), matching the size of MMMFs.

Given the value of $\overline{M}$, I recalibrate M_0^S to match the monetary injections implemented under the AMLF, implying $M_0^S = \$1,681 + \$150 = \$1,831$ billion; the value of M_2^S is set at $M_2^S = \overline{M}$ to match the fact that the AMLF was temporary. I recalibrate α as well, in order to match price stability in the economy with runs, implying $\alpha = 0.013$.

I renormalize A_1 and $\overline{K}$ to obtain $Q^* = 100$ and $P^* = 100$ in the economy with no runs. The resulting values are $\overline{K} = 8,388.19$ and $A_1 = 0.002$. I use the same value of β as in Section 6.3.2, $\beta = 0.998$.

Results. In the economy with runs, deposits are $D_0^b = \$1.113$ trillion. The difference with the calibration in Section 6.3.2 is just \$1 billion; thus, the model accounts for the same fraction of the redemptions from MMMFs. The results of the first policy analysis – the elasticities of the key endogenous variables with respect to M_0^S – have some similarities with (28) and (29), although there are also some key differences. The elasticities of velocity, v, and of the price level, P_1, are approximately the same as in Section 6.3.2, whereas the elasticity of deposits, D_0^h, is much higher (in absolute value) and the elasticity of money held by households, M_0^h, is lower:

$$\frac{dv}{dM_0^S} \times \frac{M_0^S}{v} = -0.99, \qquad \frac{dP_1}{dM_0^S} \times \frac{M_0^S}{P_1} = 0.005,$$

$$\frac{dD_0^h}{dM_0^S} \times \frac{M_0^S}{D_0^h} = -2.04, \qquad \frac{dM_0^h}{dM_0^S} \times \frac{M_0^S}{M_0^h} = 1.42.$$

In particular, the elasticity with respect to D_0^h shows that the model produces a larger reduction of deposits, in response to a temporary monetary injection. This result is confirmed by the analysis of the global effects of monetary injections. If the central bank does not inject any money ($M_0^S = \overline{M}$), the drop in the price level is much smaller (only 0.2%), but deposits are $D_0^b = \$1.298$ trillion. That is, redemptions from MMMFs ($D_0^b - D^*$) are only \$2 billion. According to this calibration, the Federal Reserve amplified the flight to liquidity by \$185 billion, much more than in the calibration in Section 6.3.2. The behavior of endogenous variables as a function of the monetary injections is qualitatively identical to that in Figure 3.

D Welfare role of banks: first best vs. bankless economy

This appendix shows that the smooth-utility model creates a role for banks in increasing welfare. To show this result, I characterize the first best that solves the social planner's problem. I consider a planner that can observe the realization of the preference shocks and is not subject to the cash-in-advance constraint. I then show that this first best is not achieved in the bankless economy, whereas it is achieved in the economy with banks when banks are not subject to runs. The effects of temporary injections on welfare are discussed briefly in Appendix C.3.

D.1 First best

Let $C_1\left(\varepsilon^H\right)$ and $C_1\left(\varepsilon^L\right)$ denote the consumption of agents hit, respectively, by preference shocks ε^H and ε^L. The problem of the planner is to choose $C_1\left(\varepsilon^H\right)$, $C_1\left(\varepsilon^L\right)$, and C_2^h for all h to maximize

$$\max_{C_1(\varepsilon^H),C_1(\varepsilon^L),\left\{C_2^h\right\}_{h\in\mathbb{H}}} \kappa\left\{\varepsilon^H u\left[C_1\left(\varepsilon^H\right)\right]\right\}+(1-\kappa)\left\{\varepsilon^L u\left[C_1\left(\varepsilon^L\right)\right]\right\}+\beta\int C_2^h dh$$

subject to the aggregate resource constraints at $t=1$ and $t=2$:

$$\begin{aligned}\kappa C_1\left(\varepsilon^H\right)+(1-\kappa)C_1\left(\varepsilon^L\right) &\leq A_1\overline{K}\\ \int C_2^h dh &\leq A_2\overline{K}+\left(1/\overline{P}_2\right)\overline{M}.\end{aligned}$$

Note that the shocks to capital do not appear in the formulation of the planner's problem, because they are idiosyncratic and, thus, do not affect the total amount of available resources.

I now characterize the solution, focusing on the time-1 allocation of consumption.[44] In the non-trivial case in which $\varepsilon^L>0$, the first-order conditions imply

$$\varepsilon^H u'\left[C_1\left(\varepsilon^H\right)\right]=\varepsilon^L u'\left[C_1\left(\varepsilon^L\right)\right]. \tag{51}$$

[44]Since the utility at $t=2$ is linear, consumption allocation among households at $t=2$ does not affect ex-ante welfare.

That is, the planner allocates consumption at $t = 1$ in order to equate the marginal utilities of consumption for patient and impatient households.

D.2 Bankless economy

I now turn to the analysis of the bankless economy in the smooth-utility model, focusing on the case $\varepsilon^L > 0$. I show that the first best is not achieved, opening up a role for banks to improve welfare.

I conjecture that households hit by the preference shock ε^H face a binding cash-in-advance constraint and, thus, spend all their money M_0^h at $t = 1$. This behavior is optimal because the nominal interest rate is positive in equilibrium (as shown later); therefore, it is not optimal to carry money that will be unspent at $t = 1$. The household problem (7) is replaced by

$$\max_{M_0^h, K_0^h, C_1(\varepsilon^H), C_1(\varepsilon^L)} \kappa \left\{ \varepsilon^H u \underbrace{\left[\frac{M_0^h}{P_1} \right]}_{= C_1(\varepsilon^H)} + \beta \frac{Q_0 K_0^h \mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2} \right\}$$
$$+ (1 - \kappa) \left\{ \varepsilon^L u \left[C_1\left(\varepsilon^L\right)\right] + \beta \frac{\left[M_0^h - P_1 C_1\left(\varepsilon^L\right)\right] + Q_0 K_0^h \mathbb{E}\left\{1 + r_2^K\left(\psi^h\right)\right\}}{\overline{P}_2} \right\} \quad (52)$$

subject to the budget constraint (8) and the cash-in-advance constraint $P_1 C_1\left(\varepsilon^L\right) \leq M_0^h$. The first-order conditions imply

$$\varepsilon^H u' \left[C_1\left(\varepsilon^H\right)\right] > \varepsilon^L u' \left[C_1\left(\varepsilon^L\right)\right]. \quad (53)$$

Thus, $C_1\left(\varepsilon^L\right)$ is greater than the social planner's optimum, and $C_1\left(\varepsilon^H\right)$ is smaller. This result arises because the cash-in-advance constraint is binding for impatient agents but is not binding for patient agents.

D.3 Economy with banks and no runs

Different from the bankless economy, the economy with banks and no runs achieves the first best. The equilibrium in the economy with banks and no runs is described by Proposition 6.1. In that equilibrium, banks pay a return on deposits that is equal to the expected return on capital, and thus the opportunity cost of holding deposits is zero – unlike the bankless economy, in which the opportunity cost of holding money is positive. As a result, no agent faces a strictly binding cash-in-advance constraint, contrary to the bankless economy.

Corollary D.1. *The equilibrium of Proposition 6.1 achieves the first best.*

Formally, the result follows from evaluating the planner's first-order condition, Equation (51), at the values of $C_1\left(\varepsilon^H\right)$ and $C_1\left(\varepsilon^L\right)$ that arise in the economy with no runs (i.e., at the values in Proposition 6.1).