Federal Reserve Bank of New York
Staff Reports

Gates, Fees, and Preemptive Runs

Marco Cipriani
Antoine Martin
Patrick McCabe
Bruno M. Parigi

Staff Report No. 670
April 2014

Gates, Fees, and Preemptive Runs
Marco Cipriani, Antoine Martin, Patrick McCabe, and Bruno M. Parigi
Federal Reserve Bank of New York Staff Reports, no. 670
April 2014
JEL classification: G21, G23, G01

Abstract

We build a model of a financial intermediary, in the tradition of Diamond and Dybvig (1983), and show that allowing the intermediary to impose redemption fees or gates in a crisis—a form of suspension of convertibility—can lead to preemptive runs. In our model, a fraction of investors (depositors) can become informed in advance about a shock to the return on the intermediary's assets. Later, the informed investors learn the realization of the shock and choose their redemption behavior based on this information. We prove two results: First, there are situations in which informed investors would wait until the uncertainty is resolved before redeeming if redemption fees or gates cannot be imposed, but those same investors would redeem preemptively if fees or gates are possible. Second, we show that for the intermediary, which maximizes the expected utility only of its own investors, imposing gates or fees can be ex post optimal. These results have important policy implications for intermediaries that are vulnerable to runs, such as money market funds, because the preemptive runs that can be caused by the possibility of gates or fees may have damaging negative externalities.

Key words: runs, gates, fees, money market funds, banks

Cipriani, Martin: Federal Reserve Bank of New York (e-mail: marco.cipriani@ny.frb.org, antoine.martin@ny.frb.org). McCabe: Board of Governors of the Federal Reserve System (e-mail: patrick.e.mccabe@frb.gov). Parigi: University of Padova (brunomaria.parigi@unipd.it).

1 Introduction

There is a longstanding view in the banking literature that the suspension of convertibility of deposits into cash can prevent self-fulfilling bank runs. However, this paper shows that the possibility that a bank might suspend convertibility can itself lead to *preemptive* runs. This result is relevant not only for an understanding of banking history and policy, but also for policy making in today's financial system. For example, recent regulatory proposals aimed at reducing the likelihood of runs on money market funds (MMFs) would give them the option to halt ("gate") redemptions or charge fees for redemptions when liquidity runs short, actions analogous to suspending the convertibility of deposits into cash at par.[1] Our results show that the option to suspend convertibility has important drawbacks; a bank or MMF with the option to suspend convertibility may become *more* fragile and vulnerable to runs.

Our focus is on preemptive runs that occur following a change in the economy's fundamentals, rather than because of a coordination failure. Hence, we build a model of a financial intermediary, in the tradition of Diamond and Dybvig (1983, hereafter "DD"). A fraction of investors become informed about an unexpected shock to the return of the investment technology, similar to that in Allen and Gale (2000). At first, informed investors only know that the return has become stochastic, but they later learn the exact realization of the shock. Our first result is that, when a gate or a fee can be imposed, informed investors may withdraw as soon as they learn about the shock, rather than waiting until they learn its exact realization. That is, they run preemptively. Our second result is that, for the intermediary that maximizes the expected utility of only its own investors, it can be ex-post optimal to impose gates or fees. To be sure, in a broader context that is beyond the scope of our model, the use of these instruments likely would not be *socially* optimal, as the intermediary would not weigh the negative externalities that might be associated with a preemptive run, such as increasing the likelihood of runs on other similar intermediaries.[2] Hence, absent commitment, the intermediary will impose a gate or fee, justifying the beliefs of informed

[1] The Securities and Exchange Commission proposed such a rule for MMFs in June 2013. Importantly, our analysis is applicable to financial intermediaries that issue liabilities that can be withdrawn or redeemed on demand but hold assets with longer maturities. These include both banks, which issue deposits to depositors, and MMFs, which issue shares to investors. Bank depositors can withdraw deposits on demand, while MMF investors can redeem shares on demand. Throughout the paper we use these terms—"depositor" and "investor," "withdraw" and "redeem"—interchangeably.

[2] See Financial Stability Oversight Council (2012) for further discussion of the large potential costs associated with runs on MMFs.

agents who run preemptively.

In the banking literature, suspension of convertibility of deposits into cash was long seen as a mechanism to prevent self-fulfilling bank runs because a suspension might eliminate the need to liquidate the investment technology. For example, in the DD model, where bank runs can occur in equilibrium as a result of a pure coordination failure, limiting withdrawals to the available liquidity is sufficient to eliminate investors' incentive to run. This suggests that the option to suspend convertibility could eliminate a centuries-old source of financial instability.

However, several papers have challenged this notion. Engineer (1989) considers a variant of the DD model with 4 periods, where investors over time bear residual uncertainty about when they need to consume. Suspension of convertibility cannot eliminate runs due to pure coordination failures in that model because some agents may want to withdraw early out of fear that suspension will prevent them from withdrawing at the date when they value consumption most. Ennis and Keister (2009, 2010) show that, absent commitment, suspension of convertibility may not prevent runs. Indeed, when a run is underway, suspension may not be ex post efficient because it would deny consumption to some impatient investors whose withdrawal requests have not been served. A policy maker who cannot commit to suspend will allow additional withdrawals, but at a discount—something akin to imposing a redemption fee. Similarly, Peck and Shell (2003) show that small modifications to the utility function make equilibrium runs possible even when convertibility can be suspended.

An important feature of these papers is that it can be very costly for investors or depositors to lose access to their funds when they need them most. This cost prevents the policy maker in Ennis and Keister (2009, 2010) from committing credibly to suspend convertibility. In the case of Engineer (1989), depositors withdraw early to reduce the likelihood that they may not be able to consume when they prefer.

This effect is also present in Freixas and Rochet (2008), who point out that suspension of convertibility is problematic when the fraction of impatient consumers is stochastic. Although it does eliminate runs, a deposit contract that allows suspension of convertibility is less efficient at risk sharing than what would be achieved by deposit insurance. The reason is that it is inefficient to set a fixed level of withdrawals that triggers suspension. If the proportion of impatient investors is larger than the fixed amount of withdrawals that the bank is willing to allow, then a run is avoided at the cost of rationing impatient investors; if, in contrast, the proportion of impatient investors is smaller than the fixed amount of withdrawals that the bank is willing to allow, then a bank run may still develop.

Our paper is the first to show that the possibility of suspending convertibility, including the imposition of gates or fees for redemptions, can *create* runs that would not otherwise occur. This contrasts with the existing literature, which focuses on whether suspension of convertibility can prevent runs. In other words, we show that rather than being part of the solution, redemption fees and gates can be part of the problem.

Our results, including our finding that restrictions on redemptions or withdrawals can be ex post optimal for the financial intermediary, provide a theoretical basis for understanding why such restrictions may have been forbidden. An intermediary's ability to impose restrictions can trigger preemptive runs with broader welfare costs that are not internalized by the intermediary or its investors. Of note, neither individual U.S. banks nor individual MMFs have had the legal option to suspend convertibility, although both types of intermediaries have done so, particularly in times of crisis. Subsequently, legislatures and courts struggled with whether and how to punish such transgressions without revoking banks' charters and forcing their liquidation after the crises had subsided (Gorton 2012).[3]

The remainder is structured as follows. Section 2 presents the model. Section 3 discusses preemptive runs. Section 4 concludes.

2 The model

2.1 Technology and preferences

The economy lasts for four dates, 0, 1, 2, and 3. It comprises a continuum of mass 1 of ex-ante identical investors and a competitive financial intermediary, such as a bank or an MMF.[4] Investors are endowed with one unit of a good that can be consumed, stored, or invested.

There are two technologies in the economy: a storage technology and an investment technology. The storage technology can be used by all agents and, as shown on line 1 of Table 1, returns one unit of the good at date $t+1$, for each unit invested at date t, $t = 0, 1, 2$. The

[3]Some of the penalties that were imposed upon U.S. banks that had suspended convertibility in the nineteenth century, such as a requirement that the bank pay a penalty rate of interest rate on liabilities until they were paid (Gorton 2012), would not be useful for all of today's financial intermediaries. For example, for a money market fund with no capital or other resources to absorb losses, a penalty rate paid to investors who had attempted to redeem shares before a suspension would reduce the value of the shares of remaining investors in the fund and increase the incentive for attempting to redeem.

[4]Equivalently, but at the cost of complicating the exposition, we could assume that many intermediaries compete for investors.

investment technology provides a return $R > 1$ at date 3 for each unit invested at date 0. The investment can be liquidated at dates 1 or 2, in which case it returns $0 \leq r < 1$ per unit invested.

Investors face a preference shock (line 2 of the table) that determines when they value consumption most. At date 1, a fraction π_1 of investors, whom we call type 1, learn that they prefer consumption at date 1. At that same time, other investors learn that they prefer consumption at either date 2 or 3, but do not learn the precise timing of their needs until date 2. At that date, a fraction π_2 of investors, whom we call type 2, learn that they prefer consumption at date 2, while the remaining $\pi_3 = 1 - \pi_1 - \pi_2$ investors, whom we call type 3, learn that they prefer consumption at date 3.

The realization of the shock to investors' preferences at dates 1 and 2 is private information. We call type-1 and type-2 investors "impatient," because they want to consume before the long-run technology matures. In contrast, type-3 investors are called "patient."

For ease of exposition, we assume that investors have logarithmic utility function, so that their expected utility at date 0 is

$$\begin{aligned} E_0 U(c_1, c_2, c_3) &= \pi_1 u(c_1 + \delta(c_2 + c_3)) + \pi_2 u(c_2 + \delta c_3) + \pi_3 u(c_3) \\ &= \pi_1 \log(c_1 + \delta(c_2 + c_3)) + \pi_2 \log(c_2 + \delta c_3) + \pi_3 \log(c_3). \end{aligned}$$

Here c_i denotes date-i consumption, with $i = 1, 2, 3$. The presence of a discount factor δ, $0 \leq \delta < 1$, implies that type-i investors prefer to consume at date t.[5]

2.2 An Unexpected Shock to the Investment Technology

We study the effect of an unexpected shock to the return of the investment technology (line 3 of the table).[6] If the shock occurs, the return of the investment technology becomes R^H with probability p and R^L with probability $1 - p$.

[5] Note that since the storage technology is available to everyone, no investor will ever consume before his preferred date. Hence, the utility function omits consumption at dates before investors' preferred consumption dates, and for simplicity, we use $u(c_1)$ also to denote the consumption of a non-type-1 investor who redeems at date 1 and stores the good for future consumption.

[6] As is standard in this literature, we think of this zero probability event as the limit case of economies in which the shock occurs with very small, but positive probability. We model this shock in a manner similar to that used by Allen and Gale (2000), who study financial contagion after the realization of a zero-probability event. In addition, Gennaioli et. al (2012) argue that investors may not take into account highly improbable risks when taking financial decisions.

We assume that the investment technology's low (high) return realization is smaller (bigger) than the storage return ($R^L < 1 < R^H$). In addition, we assume that liquidating the investment technology is costly and inefficient, even when the realization of the return is low, so $r < R^L$.[7] Hence,

$$r < R^L < 1 < R^H.$$

We also assume that $\delta R^H < 1$, so the upside potential due to the shock does not induce impatient investors to wait for a high return on the investment technology. That is, a type-i ($i = 1, 2$) investor would rather withdraw at par in period i than wait for the payoff of the investment technology.

The expected return of the technology when a shock occurs could be bigger or smaller than R. A natural assumption in studying bank runs driven by changes is fundamentals is that the shock does not increase expected returns, that is, $pR^H + (1-p)R^L \leq R$.

At date 1, a fraction q of investors, whom we call "informed," learn that the return of the investment technology is stochastic.[8] However, informed investors do not learn the exact realization of the shock until date 2. In contrast, the intermediary and the uninformed investors learn the realization at date 3. For reasons explained below, we focus on realization of q that are relatively small. In particular, we assume that:

$$q < \frac{\pi_2}{\pi_2 + \pi_3}, \tag{1}$$

and

$$q < r. \tag{2}$$

At least for some intermediaries, assumption 2 is not very constraining from an empirical standpoint. For example, liquidation-related losses $(1 - r)$ are likely to be quite small for MMFs, which historically have experienced losses (including credit losses) that are only a few percent of assets, so r would probably exceed 0.9. Moroever, a relatively high estimate of q might be 0.6, based on the largest fractions of shares redeemed from individual institutional prime MMFs in the week following the bankruptcy of Lehman Brothers in September 2008.[9]

[7] For example, amidst a run on the intermediary, sale of the asset before it matures might only be possible at a fire-sale prices.

[8] Note that it makes no difference conceptually whether q is a fraction of all investors or only of type 2 and 3 investors, since type-1 investors do not alter their behavior after the shock. We also assume that type-2 and type-3 investors are equally likely to be informed.

[9] Based on data from iMoneyNet and authors' calculations.

2.3 Redemption Fees and Gates

Our goal is to study whether a financial intermediary's ability to impose restrictions on withdrawals can cause instability. We consider two types of restrictions. Both restrictions are liquidity contingent, and both aim to limit or prevent the inefficient liquidation of the investment technology. In particular, we consider:

(A) A liquidity-contingent ("standby") gate, which halts redemptions (withdrawals) when the intermediary's liquid assets are depleted. That is, the gate suspends convertibility of the claims against the financial intermediary into cash.

(B) A liquidity-contingent ("standby") fee, that is, a rule that taxes redemptions (withdrawals) when the intermediary's liquidity is depleted. The intermediary collects the fee on any redemptions in excess of its liquid assets and holds the proceeds in the storage technology for subsequent distribution to remaining investors.

Note that both gates and fees are triggered by the depletion of the liquid assets held by the intermediary. However, the gate postpones a withdrawal, whereas a fee reduces the amount that can be withdrawn. In addition, a fee redistributes resources from investors who withdraw before the investment technology pays off to those who redeem at date 3.

In the model, we assume that the sequential service constraint holds. That is, investors' withdrawals are processed on a first-come first-served basis until a gate or a fee is imposed or the intermediary's resources are exhausted.

2.4 The Optimal Contract

As is standard in this literature, the intermediary is competitive—that is, subject to a zero-profit condition—and chooses the contract that maximizes the expected utility of its investors, subject to the available technologies. Since the shock is unexpected at date 0, it does not affect the optimal contract.

At date 0, investors deposit their endowments with the intermediary in order to diversify their liquidity risk. The intermediary invests π_3 in the investment technology and stores $\pi_1 + \pi_2$ for type-1 and 2 investors. As a result, the optimal contract, denoted by c^*, is given by:[10]

$$c_1^* = 1, c_2^* = 1, c_3^* = R.$$

[10] The optimal contract follows from the Euler equation of consumption,
$U'(c_1^*) = U'(c_2^*) = RU'(c_3^*)$,
which can be easily generalized for any well-behaved utility function.

The optimal contract is summarized on line 4 of the table, along with equilibrium investors' choices.

3 Runs

In this paper, we focus on runs that occur following a change in the economy's fundamentals, rather than because of a coordination failure. This approach is consistent with evidence described by Gorton (1985), who documents that historically crises have often occurred in economies with weak fundamentals.

Our aim is to show that a bank's or MMF's ability to impose restrictions on withdrawals puts it at risk of preemptive runs. By a preemptive run, we mean a run that occurs in anticipation of the imposition of the restriction. We consider environments in which, if the intermediary *cannot* impose restrictions on withdrawals, informed investors wait until uncertainty is resolved, at date 2, before deciding whether to withdraw. In contrast, if the intermediary *can* impose restrictions on withdrawals, a preemptive run may occur with informed investors withdrawing at date 1, before they know how the uncertainty with be resolved.

In our model, as in Engineer (1989), sunspot runs can occur even if the intermediary imposes fees or suspends convertibility; that is, restrictions to withdrawals do not eliminate a sunspot run. This, however, is not the focus of our paper. Instead, our analysis shows that, even without sunspots, the intermediary's ability to restrict redemptions (withdrawals) may cause preemptive runs.

3.1 No-run Condition

First, we consider an economy where the intermediary cannot impose a fee or a gate on redemptions (withdrawals). We show that, when gates or fees cannot be imposed, informed investors decide not to withdraw at date 1 once they learn about the shock. That is, preemptive runs do not occur.

Informed investors will not run preemptively at date 1 when:

$$u(\widehat{c}_1) < p[\gamma u(\widehat{c}_2^H) + (1-\gamma)u(\widehat{c}_3^H)] + (1-p)u(\widehat{c}_2^L), \tag{3}$$

where $\widehat{c}$ denotes the consumption levels after the shock has occurred, and γ is the share of type-2 investors among non-type-1 investors, that is, $\gamma \equiv \frac{\pi_2}{\pi_2+\pi_3}$.

The LHS of expression (3) represents the utility of an informed investor who withdraws at date 1. The RHS represents the investor's expected utility if he waits for uncertainty about the return on the investment technology to be resolved at date 2 before making a withdrawal decision. With probability p, the return of the investment technology is R^H, in which case the informed investor consumes $\widehat{c}_2^H$ if he is of type 2 and $\widehat{c}_3^H$ if he is of type 3. Otherwise, the return of the investment technology is R^L, in which case the informed investor withdraws at date 2 and consumes $\widehat{c}_2^L$.

Several observations are in order. First, $\widehat{c}_1 = 1$ since we are considering a case where there is no run at date 1, and thus a deviating informed investor obtains 1 at date 1. Second, $\widehat{c}_2^H = 1$ since only type-2 investors withdraw at date 2 when the return is high; type-3 investors wait until date 3 because $R^H > 1$. Third, $\widehat{c}_3^H = R^H$ because only type-3 investors withdraw at date 3 when the return is high; type-2 investors withdraw at date 2 even if they learn that the technology returns R^H, since $\delta R^H < 1$.

Fourth, informed type-3 investors have an incentive to withdraw at date 2 when the return of the long-run technology is low, since $R^L < 1$. Nevertheless, $\widehat{c}_2^L = 1$ even if informed type-3 investors withdraw at date 2 along with type-2 investors. This follows from the assumption that $q < r$, which implies

$$\pi_2 + \pi_3 r > \pi_2 + \pi_3 q. \tag{4}$$

Here, the LHS is the value of the assets at date 2, and the RHS is the volume of withdrawals at date 2 in the low-return state. Even if all informed investors withdraw at date 2, the financial intermediary remains solvent, although it has to liquidate of some of the investment technology to pay all of the withdrawing investors.

Equation (3) always holds since

$$u(1) < p[\gamma u(1) + (1-\gamma)u(R^H)] + (1-p)u(1). \tag{5}$$

Thus, absent restrictions on withdrawals, informed investors do not run preemptively. Instead, they wait for the resolution of uncertainty at date 2 to make their withdrawal decisions. Because the portion of informed investors is small enough ($q < r$), the intermediary has enough resources to pay them par at date 2 in the low-return state. Hence, preemptive runs never occur in this economy.

3.2 The Gate

In this section, we show that the *possibility* of imposing a gate can create a preemptive run at date 1. We then show that such a gate can be ex-post optimal for the intermediary, which maximizes the expected utility of its investors. Hence, absent commitment, informed investors anticipate that the intermediary will halt redemptions and choose to run preemptively.

3.2.1 Preemptive Runs with the Gate

As discussed above, the intermediary imposes the gate once cumulative attempted withdrawals (redemptions) reach total liquid assets. We focus on the effect of a gate at time 2, since, as will be explained below, the intermediary never imposes a gate at date 1. Assuming that there was no preemptive run at date 1, a gate would be imposed at date 2 if attempted withdrawals exceed π_2. The gate *delays* further redemptions until the investment technology matures at date 3.

A preemptive run at date 1 is optimal for the informed investors if:

$$u(\tilde{c}_1) > p[\gamma u(\tilde{c}_2^H)+(1-\gamma)u(\tilde{c}_3^H)]+(1-p)\left[\gamma V_2^L(\circ) + (1-\gamma)V_3^L(\circ)\right]. \quad (6)$$

Here, $\tilde{c}$ refers to consumption levels when the gate is imposed, and $V_2^L(\circ)$ and $V_3^L(\circ)$ are the type-2 and type-3 investors' expected utilities in the low-return state if there is no preemptive run. As is the case in expression (3), the LHS of expression (6) represents the utility of an investor who withdraws at date 1, whereas the RHS represents the investor's expected utility if he waits until date 2 before making a redemption decision.

We make several observations. First, $\tilde{c}_1 = 1$. This follows from the assumption that $q < r$, which ensures that, even if a preemptive run occurs, the intermediary's available resources at date 1 exceed all redemption requests by type-1 and informed type-2 and type-3 investors, that is:

$$\pi_1 + \pi_2 + \pi_3 r > \pi_1 + q(\pi_2 + \pi_3). \quad (7)$$

In addition, assumption 1 ($q < \frac{\pi_2}{\pi_2+\pi_3}$) guarantees that the intermediary's available liquidity exceeds redemptions in a preemptive run, that is,

$$\pi_1 + \pi_2 > \pi_1 + q(\pi_2 + \pi_3). \quad (8)$$

Therefore, even if the intermediary has the option to impose a gate, it will not do so at date 1.

Second, as in the previous section, $\tilde{c}_2^H = 1$ and $\tilde{c}_3^H = R^H$, since when the state is high type-2 and type-3 investors have an incentive to withdraw at their preferred time (dates 2 and 3, respectively). Third, an informed type-3 investor will try to withdraw at date 2 if R^L is realized. The investor's expected payoff in this situation is

$$V_3^L(\circ) \equiv u(1)\frac{\pi_2}{\pi_2 + q\pi_3} + u(R^L)(1 - \frac{\pi_2}{\pi_2 + q\pi_3}) > u(R^L), \quad (9)$$

which is a weighted average of the investor's utility under two possible outcomes. Under the sequential service constraint, the intermediary pays fraction $\frac{\pi_2}{\pi_2+\pi_3 q}$ of type-3 informed investors at date 2 before the gate is imposed, and their utility is $u(1)$. This fraction is the ratio of available liquid assets at date 2, π_2, to the total volume of attempted redemptions, $\pi_2 + \pi_3 q$, which includes attempted redemptions by all type-2 investors and informed type-3 investors. A fraction $1 - \frac{\pi_2}{\pi_2+q\pi_3}$ of investors who want to withdraw at date 2 cannot do so because of the gate and receive $u(R^L)$. Since $R^L < 1$, $u(1) > V_3^L(\circ)$.

Fourth, the expected utility of investors of type 2 is given by

$$V_2^L(\circ) \equiv u(1)\frac{\pi_2}{\pi_2 + q\pi_3} + u(\delta R^L)(1 - \frac{\pi_2}{\pi_2 + q\pi_3}).$$

Since investors of type 2 who cannot withdraw at date 2 cannot consume at the date at which they value consumption most, $V_2^L(\circ) < V_3^L(\circ) < u(1)$.

Hence, condition (6) holds whenever:

$$u(1) > p[\gamma u(1) + (1-\gamma)u(R^H)] + (1-p)V_3^L(\circ). \tag{10}$$

Since $V_3^L(\circ) < u(1)$, there is a value of p small enough that (10) holds for any R^H and R^L. In other words, if the shock brings sufficiently bad news about fundamentals (a low value of p), a preemptive run will occur. Note that the preemptive run is not observed by uninformed investors, and hence they do not modify their behavior.

3.2.2 Ex-post Optimality of a Gate

So far, we have shown that preemptive runs occur *only* if the intermediary can gate withdrawals. In this section, we prove that for an intermediary that seeks to maximize the expected utility of (only) its investors, imposing a gate can be ex-post optimal.[11] This result is important because an intermediary that has the ability to suspend convertibility may like to commit not to do so in order to avoid preemptive runs. However, if it cannot commit, it will suspend convertibility at date 2. In doing so, it would validate the beliefs of informed investors who had anticipated that a suspension would occur and hence had run preemptively.

This suggests an explanation for why suspension of convertibility historically was prohibited for banks, at least in the U.S., and only occurred in very rare cases. Banks with a legal option to suspend likely

[11] As noted earlier, our model does not address the potential social costs associated with preemptive runs, such as the possibility that they would make runs on other intermediaries more likely and lead to a reduction in credit supply. In a broader context, an intermediary that maximizes the expected utility of its own investors may impose a gate, even if consideration of externalities would show that this choice is not *socially* optimal.

would have been expected to do so, and thus preemptive runs might have been common. Similarly, our result highlights a concern about providing MMFs with the option to impose gates or fees.

Our analysis can be limited to the case in which the return to the investment technology is R^L; if it is R^H, informed type-3 investors wait until date 3 to redeem, so there is no reason for the intermediary to impose a gate at date 2. Also, the optimality of imposing a gate at date 2, after type-1 investors have already withdrawn, depends only on the expected utility of type-2 and type-3 investors. If the intermediary does not suspend convertibility, their expected utility in the low-return state is:

$$\pi_2 \log(1) + q\pi_3 \log(1) + (1-q)\pi_3 \log((1-\frac{q}{r})(\frac{1}{1-q})R^L)). \tag{11}$$

The first two terms are the utility of type-2 and informed type-3 investors, respectively, and the third term is the utility of uninformed type-3 investors. This last term reflects the fact that $\frac{q}{r}\pi_3$ of the investment technology has been liquidated at date 2 to satisfy informed type-3 investors.

If, instead, convertibility is suspended, only a measure π_2 of investors consume $c_2 = 1$ at date 2. A measure $\pi_2 \frac{\pi_2}{\pi_2+q\pi_3}$ of them are of type 2, whereas a measure $\pi_2 \frac{q\pi_3}{\pi_2+q\pi_3}$ of them are type 3. In addition, a measure π_3 of investors consume R^L at date 3. A measure $\pi_2(1-\frac{\pi_2}{\pi_2+q\pi_3}) = \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3}$ of them are of type 2 whereas a measure $(1-q)\pi_3 + \left(q\pi_3 - \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3}\right) = \pi_3 - \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3}$ of them are of type 3. Therefore, the expected utility of a type-2 or type-3 investor can be written

$$\pi_2 \log(1) + \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3} \log \delta R^L + \left(\pi_3 - \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3}\right) \log R^L \tag{12}$$

Thus, suspension of convertibility is ex-post optimal for the intermediary whenever (11) < (12). This condition can be written

$$\begin{aligned} q\pi_3 \log(1) + (1-q)\pi_3 \log((1-\frac{q}{r})(\frac{1}{1-q})R^L)) \\ < \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3} \log \delta R^L + \left(\pi_3 - \frac{\pi_2 q\pi_3}{\pi_2+q\pi_3}\right) \log R^L \end{aligned}$$

$$(1-q)\log\left[\left(\frac{1-\frac{q}{r}}{1-q}\right) R^L\right] < \frac{\pi_2 q}{\pi_2+q\pi_3} \log \delta R^L + \left(1 - \frac{\pi_2 q}{\pi_2+q\pi_3}\right) \log R^L.$$

Gating will be optimal, for example, when q approaches r since the LHS tends to minus infinity; in this circumstance, the exess withdrawal at time 2 may be high relative to the cost of liquidating the investment technology, which makes the use of a gate to prevent liquidation more attractive. Moreover, large values of δ (which cause the RHS to approach to $\log(R^L)$) reduce the cost of delaying the consumption of type-2 investors and make gating more appealing. In such cases, suspension of convertibility is ex-post optimal for the intermediary. Indeed, for an intermediary that can impose a gate, a policy of not doing so at date 2 is not time consistent. Informed investors anticipate that convertibility will be suspended and run preemptively.

3.3 The Fee

In this section, we examine the possibility of a fee. As in the previous section, we first consider an economy where the intermediary cannot impose a fee (i.e., it can commit not to impose the fee) and then one in which it *can* impose the fee. We obtain results similar to those for the gate: when the financial intermediary cannot impose a fee, a preemptive run will never occur, and informed investors will wait until dates 2 and 3 to withdraw.

The fee that the intermediary can impose is liquidity-contingent; it can be put in place when the intermediary's liquidity is depleted. The fee, $\phi \in [0, 1)$, is a share of the value of withdrawals, so if a fee is triggered, an investor withdrawing 1 receives an amount $1 - \phi > 0$. The intermediary collects the fee on any redemptions in excess of its liquid assets and holds the proceeds in the storage technology for subsequent distribution to remaining investors.

3.3.1 Preemptive Runs with Fees

In this section, we show that the *possibility* of a fee on withdrawals can create a preemptive run at date 1. In the next section, we will show that such a fee is ex-post optimal for the intermediary. Hence, absent commitment, informed investors anticipate that the intermediary will impose a fee and choose to run preemptively.

A preemptive run at date 1 is optimal for the informed investors if

$$u(\tilde{c}_1) > p[\gamma u(\tilde{c}_2^H) + (1-\gamma)u(\tilde{c}_3^H)] + (1-p)\left[\gamma V_2^L(\circ) + (1-\gamma)V_3^L(\circ)\right]. \tag{13}$$

The expression (13) is analogous to (6), but here, $\tilde{c}$ refers to consumption levels when a fee is imposed. Again, $V_2^L(\circ)$ and $V_3^L(\circ)$ are the type-2 and type-3 investors' expected utilities in the low-return state if there is no preemptive run. The LHS of expression (13) represents the utility of an investor who withdraws at date 1, whereas the RHS represents the investor's expected utility if he waits one more period before making a decision.

Several observations are in order. First, $\tilde{c}_1 = 1$, for the reasons discussed in section 3.2.1. That is, even with a preemptive run, redemptions at date 1 do not exceed the intermediary's total resources or its available liquidity. Thus, the intermediary never imposes a fee at date 1. Second, as discussed in section 3.2.1, $\tilde{c}_2^H = 1$ and $\tilde{c}_3^H = R^H$.

For condition (13) to hold, $V_2^L(\circ)$ and $V_3^L(\circ)$, the expected utility of type-2 and type-3 informed agents in the low-return state, needs to be less than $u(\tilde{c}_1) = u(1)$. Note that $V_3^L(\circ) \geq V_2^L(\circ)$, as type-2 investors discount consumption at date 3. Therefore, it is enough to show that $V_3^L(\circ) < u(1)$. The term $V_3^L(\circ)$ is the weighted average of the utilities from three possible levels of consumption: (a) consumption of 1, the payoff at time 2 if an informed type-3 agent redeems before the fee is imposed; (b) consumption of $1 - \phi$, the payoff at time 2 if an informed type-3 agent withdraws after a fee is imposed; (c) the amount paid to any agent withdrawing at date 3 (and hence, the payoff of an informed type-3 agent who decides, after a fee has been imposed, to postpone redeeming until time 3). This last amount must be smaller than 1. If it were not, all type-3 agents would withdraw at 3, and no fee would be imposed; in this case, there would be no fee paid to the intermediary and the payout at time 3 would be $R_L < 1$; a contradiction. Therefore, $u(1) > V_3^L(\circ) \geq V_2^L(\circ)$.[12]

For a p small enough, there will be a preemptive run, since:

$$u(1) > p[\gamma u(1) + (1-\gamma)u(R^H)] + (1-p)\left[\gamma V_2^L(\circ) + (1-\gamma)V_3^L(\circ)\right].$$

Note that the preemptive run may occur for any level of the fee. Since any fee pushes date-2 consumption below 1, there will always be a p small enough that agents have an incentive to run preemptively.

3.3.2 Ex-post Optimality of a Fee

In this section, we want to show that an intermediary that has the ability to impose a fee at date 2 will find it ex-post optimal to do so. Let us assume that, at date 2, R^L is realized. Consider a ϕ small enough so as not to alter the behavior of type-3 informed investors withdrawing at date 2. That is, informed type-3 investors who learn that the investment technology yields R^L decide to withdraw at date 2 (obtaining $1-\phi$) even if they are assessed the fee.

Imposing such a fee has two effects. First, it reduces the amount of investment technology being liquidated, since a fraction of investors

[12] Note that in order to prove the existence of preemptive runs, we do not need to compute agents' payoff for a given fee. The argument also does not rely on the specific level of the optimal fee that the intermediary would choose to impose.

withdrawing at date 2 obtain $1-\phi$ instead of 1. Since $r < R^L$, such a reduction increases the overall amount of resources to be distributed across investors. Second, the fee redistributes resources from investors withdrawing at date 2 (who obtain either 1 or an amount $1-\phi$) to type-3 uninformed investors (who obtain an amount strictly smaller than $1-\phi$).[13] Since the investors' utility function is concave, if the intermediary maximizes the sum of the welfare of all its investors (that is, if it cares equally about informed and uninformed investors), it will prefer to impose such a fee to not imposing any fee. That is, a policy of not imposing a fee at date 2 is not time consistent; an intermediary with the ability to charge such a fee will do so.

4 Conclusion

We build a model in the tradition of Diamond and Dybvig (1983) with a shock to the investment technology occurring unexpectedly at date 1. Some investors are informed and can choose to redeem (withdraw) immediately, or wait until they learn the exact realization of the shock, at date 2. A preemptive run occurs when informed investors withdraw at date 1 instead.

We prove two results: First, we show that there can be preemptive runs that occur *only* because an intermediary has the ability to impose "standby" (liquidity-contingent) gates or fees. Second, we show that for an intermediary that maximizes the expected utility of its own investors, imposing a gate or fee can be ex-post optimal. Hence, for an intermediary that can restrict redemptions in a crisis, a policy of not imposing such restrictions may be time-inconsistent. The financial intermediary might like to commit not to restrict redemptions, so that preemptive runs would not occur. Absent a means of ensuring commitment, however, the intermediary will find it optimal to suspend, confirming the beliefs of informed investors who withdrew preemptively.

Our results have broader policy significance. Rules that provide intermediaries, such as MMFs, the ability to restrict redemptions when liquidity falls short may threaten financial stability by setting up the possibility of preemptive runs. Much of the wider policy significance of that risk is beyond the scope of this paper, since our model does not incorporate the large negative externalities associated with runs on financial institutions, including MMFs. But one notable concern, given the similarity of MMF portfolios, is that a preemptive run on one fund might cause investors in other funds to reassess whether risks in their

[13] This is true since otherwise informed type-3 investors would decide not to pay the fee and wait until time 3 to obtain a higher payout.

funds are indeed vanishingly small.

5 References

Allen, Franklin, and Douglas Gale (2000), "Financial Contagion" *Journal of Political Economy* 108:1 (February), 1-33.

Diamond, Douglas W. and Philip H. Dybvig (1983), "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* 91:3 (June), 401-419.

Ennis, Huberto M. and Todd Keister (2009), "Bank Runs and Institutions: The Perils of Intervention," *American Economics Review* 99:4 (September), 1588-1607.

Ennis, Huberto M. and Todd Keister, (2010), "Banking Panics and Policy Responses," *Journal of Monetary Economics* 57:4 (May), 404-419.

Engineer, Merwan (1989), "Bank Runs and the Suspension of Deposit Convertibility," *Journal of Monetary Economics* 24:3 (November), 443-454.

Financial Stability Oversight Council (2012), "Proposed Recommendations Regarding Money Market Mutual Fund Reform," *Federal Register* 77:223 (November 19) 69455-83.

Freixas, Xavier and Jean-Charles Rochet (2008), *Microeconomics of Banking*, Cambridge, MA: MIT Press.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny (2012), "Neglected Risks, Financial Innovation, and Financial Fragility," *Journal of Financial Economics* 104:3 (June), 452-468.

Gorton, Gary (1985), "Bank Suspension of Convertibility" *Journal of Monetary Economics* 15:2 (March), 177-93.

Gorton, Gary (2012), *Misunderstanding Financial Crises: Why We Don't See Them Coming*, Oxford: Oxford University Press.

Peck, James and Karl Shell (2003), "Equilibrium Bank Runs," *Journal of Political Economy* 111:1 (February), 103-123.

Table 1. Model timeline.

Date (t)	t = 0	t = 1	t = 2	t = 3
1. Technology				
a. Storage: opportunity at any date t=0,1,2. Payoffs at t+1:	Store 1	1	1	1
b. Investment: opportunity only at date 0. Payoffs:	Invest 1	r<1	r<1	R
2. Preference shock		Type-1 investors (measure π_1) learn privately that they prefer to consume at date 1; residual uncertainty remains for types 2 and 3.	Type-2 and type-3 investors (measures π_2 and π_3, respectively) learn privately that they prefer to consume at dates 2 and 3, respectively.	
3. Technology shock		Informed investors (fraction q) learn privately that shock has occurred: $R=R^H$ (R^L) with probability p (1-p)	Informed agents learn realization the shock.	Uninformed agents and financial intermediary (FI) learn realization of R.
4. Optimal contract and equilibrium when R>1.	All investors deposit 1. FI invests π_3 in the investment technology and $\pi_1 + \pi_2$ in the storage technology.	FI's liquid assets are $\pi_1 + \pi_2$. Redeeming investors receive 1. Each type-1 investor redeems, so redemptions total π_1.	FI's remaining liquid assets are π_2. Redeeming investors receive 1. Each type-2 investor redeems, so redemptions total π_2.	Investment technology matures with return R. Redeeming investors receive R.
5. Gate: FI halts redemptions (withdrawals) when liquid assets depleted.		Gate imposed if attempted redemptions exceed FI's liquid assets, $\pi_1 + \pi_2$.	Gate imposed if attempted redemptions exceed FI's remaining liquid assets, π_2.	
6. Fee: FI taxes redemptions (withdrawals) at rate ϕ when liquid assets depleted.		Fee imposed if attempted redemptions exceed FI's liquid assets, $\pi_1 + \pi_2$.	Fee imposed if attempted redemptions exceed FI's remaining liquid assets, π_2.	