Financial Intermediation and Fire Sales with Liquidity-Risk Pricing *

Yuteng Cheng Bank of Canada

Roberto Robatto University of Wisconsin-Madison

November 18, 2022

Abstract

We provide a novel theory of fire sales based on *liquidity-risk pricing.* Fire-sale prices of illiquid financial assets arise when financial intermediaries are in distress and do not provide investors with full insurance against liquidity risk. This liquidity-risk pricing does not rely on standard mechanisms used in the literature such as cash-in-the-market pricing, second-best use, or asymmetric information. In particular, a single friction—liquidity risk—plays a dual function in providing a role for financial intermediaries and generating fire-sale prices. Our framework encompasses several classes of models, and we provide two applications. First, we embed our mechanism in a standard bank runs model. Second, we consider a model in which investors trade with dealers in OTC markets. Policy implications differ substantially from those of other fire-sales models in the literature: ex-ante interventions that alter investments in liquid and illiquid assets—such as liquidity requirements—are counterproductive.

*Cheng: ycheng@bank-banque-canada.ca. Robatto: robatto@wisc.edu. A previous version of this paper circulated under the title "Bank Runs and Endogenous Fire Sales with Consumption-Based Asset Pricing." We thank Pablo Kurlat, William Diamond, and Alireza Tahbaz-Salehi for helpful conversations and suggestions. The views expressed herein are those of the authors and not necessarily those of the Bank of Canada.

1 Introduction

During several episodes of financial crises, many financial assets are traded at so-called fire-sale prices, that is, prices that are somewhat lower than the fundamental value (Merrill et al. 2021). Fire-sale prices are often problematic, as they might have a negative impact on the balance sheet of leveraged financial and they might precipitate runs on banks and other bank-like institutions such as money market mutual funds (Schmidt, Timmermann, and Wermers 2016). Because of the importance of these events, academics and policymakers have analyzed them in great details, and regulators have introduced several policies such as liquidity requirements on banks and money market funds aimed at reducing these events and their spillovers on the real economy.

Most theories of fire sales rely on either cash-in-the-market pricing (Allen and Gale, 1998), second-best use (Shleifer and Vishny, 1992; Kiyotaki and Moore, 1997; Lorenzoni, 2008), or asymmetric information (Kurlat, 2016). These theories differ from one to another, but they all share a common feature. That is, while asset prices are in general equal to expected discounted cash flows, fire-sale prices in these theories are typically the result of the expectation part only, with little or no role played by fluctuations in discount rates (i.e., investors' marginal utilities). Indeed, many of these theories can be derived in frameworks in which marginal investors have linear utility (Allen and Gale, 1998; Kurlat, 2021). Yet, this approach is in stark contrast with modern asset pricing, which emphasizes the importance of fluctuations in discount rates to explain movements in asset prices (Cochrane, 2011).

This paper presents a novel theory of fire-sale pricing that fills this gap and studies the resulting policy implications. We characterize a large class of models in which a fire-sale price arises when financial intermediaries do not provide investors with full insurance against liquidity risk, and investors are active traders in financial markets. When financial intermediaries are malfunctioning, investors are exposed to liquidity risk, resulting in a distortion to their stochastic discount factor—that prices securities in equilibrium. In particular, exposure to liquidity risk reduces investors' demand for illiquid assets, and as long as asset supplies are not fully elastic, produces a low fire-sale price. We refer to this mechanism as *liquidity-risk pricing*. Importantly, a single friction—liquidity risk—plays a dual function in providing a role for financial intermediaries and generating fire-sale prices.

We first present our theory in a general framework that encompasses several classes of models used in the financial intermediation literature. The framework is purposefully general and includes only the key elements of the environment and the key features of the equilibrium that are relevant to understand fire-sale prices driven by the liquidity-risk pricing mechanism and derive the key policy results. We then provide two applications: (i) a banking model in which financial intermediaries are subject to runs, and (ii) a model of over-the-counter (OTC) trades in which investors trade with dealers in decentralized markets.

Our theory of fire sales based on liquidity-risk pricing has important policy implications. In response to the 2008 and COVID- 19 crises, several new regulations have been imposed to limit the risk and negative effects of fire sales, such as liquidity requirements on banks and mutual funds. This has been in part motivated by the theoretical fire-sale literature, that typically finds externalities associated with fire-sales (Dávila and Korinek 2017; Lorenzoni 2008; Kurlat 2021). Our policy implications, however, are very different. In our framework, we show that ex-ante investments in liquid and illiquid assets are efficient, in the sense that a regulator that internalizes the effects on prices would make the same decisions as price-takers financial agents. Hence, liquidity requirements are counterproductive, in the sense that they reduce welfare—by forcing investments in liquidity, this policy reduces investments in more productive long-term projects. The result follows directly from the way fire sales work in our framework. In contrast to, say, models of fire sales based on borrowing constraints and second-best use (e.g., Lorenzoni (2008)), the fire-sale price here does not enter into any binding borrowing constraint. Hence, the main channel that creates inefficiency and amplifications is absent here. However—and stepping a bit outside the model—interventions that increase liquidity in crisis times might be helpful to put upward pressure on the price of long-term assets and, ultimately, redirect initial investments toward highly productive projects.

The relevance of our policy recommendations depends on whether our framework is a good description of actual fire sales. In this sense, we note that a key element of our model is that investors can hold not only deposits (or intermediaries' debt more generally) but also the same securities that intermediaries have in their balance sheets. Thus, the framework could apply to e.g. certain money markets and bond mutual funds, such as those that target institutional investors. Such investors might have the ability to not only hold the funds' shares but also purchase e.g. commercial paper or corporate bonds. Hence, our theory suggests that liquidity regulations imposed on these entities reduces welfare.

Let us now provide more details about our model. In our general framework, we consider a discrete-time setting and focus on three periods: t, $t+1$, and $t+2$. At time t, investors have access to a centralized market in which they can trade a short asset (i.e., liquidity) and a long-term asset that pays off at $t+2$. At $t+1$, investors are subject to liquidity shocks—modeled as preference shocks as in standard banking and OTC models (Diamond and Dybvig, 1983; Weill, 2020). Crucially, when the liquidity risk is materialized, investors have limited market access, and each of them can only meet a financial intermediary to e.g. trade in an OTC market or withdraw deposits. When disruption in financial markets impairs intermediaries' ability to trade or pay out deposits, investors are exposed to liquidity risk. Anticipating this possibility, investors tilt their time-t demand for investments toward liquidity and away from the long-term asset. As long as the supply of liquidity and of the long-term asset are somewhat rigid, the price at which the long-term asset trades at time t drops to a fire-sale level in comparison to normal times (i.e., in comparison to a scenario in which intermediaries are well-functioning and provide better insurance against liquidity risk). Formally, the exposure to liquidity risk tilts the pricing kernel that investors use to price the long-term asset, reducing the equilibrium price.

In our first application, we consider fire sales in a model of panic-based runs. We extend Diamond and Dybvig (1983) by introducing a centralized market in which banks and depositors can trade liquidity and a long-term asset before the realization of all the liquidity risk, and we maintain the assumption that no trade is possible after. In the bad equilibrium, depositors run, and banks liquidate their long-term investments by selling them in the centralized market at a fire-sale price. Crucially, depositors are on the other side of the trade—they purchase the long-term assets that banks sell. To understand why a fire-sale price arises, note that the runs are associated with bank failures, leaving depositors with no insurance against liquidity risk. Hence, depositors are willing to pay a low price for the the long-term asset, as they tilt their demand toward liquidity. The low price leads to banks' insolvencies, making runs a self-fulfilling event. In contrast, in the good equilibrium, the liquidity risk is fully offset by intermediaries through demand deposits, leading to a high price for the long-term asset and a safe and stable banking sector.

In our second application, we study a simple model of OTC trades. The market structure is somewhat similar to that of the bank runs application—investors have first access to a centralized market where they can trade liquidity and a long-term asset, and are then exposed to liquidity shocks in the form of heterogeneous consumption needs. As discussed by Weill (2020), this type of liquidity shocks—when embedded in a model of OTC trade—can translate into the more commonly used indirect utility for financial assets. In our model, after the realization of the liquidity shocks, investors can meet dealers and engage in OTC trade to either get more liquidity (if they are subject to a liquidity shock) or purchase more of the long-term asset (if they are not subject to the liquidity shock). We consider two extreme cases—one in which dealers are met with certainty and the terms of trade generate little spreads, and another one in which OTC trades are exogenously shut down—the latter case could arise if dealers are not willing to engage in any trade because of severe distress. When OTC trades are shut down, investors preemptively try to tilt their portfolio toward liquidity and away from the long-term asset before the realization of the liquidity risk. However, as long as asset supplies are somewhat rigid, the outcome is a reduction in the price of the long-term asset.

2 General framework

We begin by providing a general framework in which we show how investors' liquidity risk generates fire sales when financial intermediaries are not well-functioning. The framework is purposefully general, and Sections 3 and 4 provide applications to bank runs and OTC markets in which we provide the full details of the environment and fully characterize the equilibrium. The objective of this section is to sketch only the key elements of the environment that are common to the various applications that we will analyze, as well as the features of the equilibrium related to asset pricing and, in particular, fire sales. Section 2.1 present the environment and the main result. Section 2.2 characterizes the efficiency properties of the equilibrium and show that ex-ante regulatory intervention that alter investments in liquid and illiquid assets—such as liquidity requirements—are counterproductive in our environment.

2.1 Environment and main result

Consider an economy where time is discrete. The analysis here applies to both a finite or infinite horizon model, with our focus being on period t, $t+1$, and $t+2$.

Investors have preferences

$$U\left(\{c_{t+1}, c_{t+2}, ...\}; \varepsilon\right) \tag{1}$$

where ε is a preference shock with distribution $F(\varepsilon)$ that affects the marginal utility of consumption at time $t+1$:

$$\frac{\partial U\left(\{c_{t+1}, c_{t+2}, ...\}; \varepsilon'\right)}{\partial c_{t+1}} > \frac{\partial U\left(\{c_{t+1}, c_{t+2}, ...\}; \varepsilon\right)}{\partial c_{t+1}} \quad \text{for } \varepsilon' > \varepsilon.$$

In words, higher values of ε are associated with higher marginal utility at time $t+1$. Thus, ε captures standard liquidity shocks that are used to motivate trades in models of OTC markets or demand-deposit contracts in banking models.

We assume that the utility function $U(\cdot)$ in (1) is strictly concave in c_{t+1}, and that $\partial U(\cdot)/\partial c_{t+1}$ is convex in c_{t+1}. We also impose the sufficient condition that $U(\cdot)$ exhibits constant marginal utility from time-2 consumption:

$$\frac{\partial U\left(\{c_{t+1}, c_{t+2}, ...\};\varepsilon\right)}{\partial c_{t+2}} = \beta, \tag{2}$$

where $0 < \beta < 1$. That is, $U(\cdot)$ is linear in c_{t+2}, and β denotes the discount factor. This is again in line with several OTC models as well as infinite-horizon models that use quasi-linear preference to obtain a linear continuation utility (Lagos and Wright, 2005). The linearity assumption can be relaxed, but it is important that, in equilibrium, investors' marginal utilities at $t+2$ do not strongly co-move with marginal utilities at $t+1$.

From a planner perspective, a Pareto-optimal allocation equalizes marginal utilities at $t+1$, that is,

$$\frac{\partial U\left(\{c_{t+1}(\varepsilon), c_{t+2}, ...\};\varepsilon\right)}{\partial c_{t+1}} = \frac{\partial U\left(\{c_{t+1}(\varepsilon'), c_{t+2}, ...\};\varepsilon\right)}{\partial c_{t+1}} \quad \text{for all } \varepsilon,\, \varepsilon'. \tag{3}$$

Without loss of generality, we normalize such marginal utilities to one—an assumption that simplifies the exposition.

The market structure plays a crucial role in our results. At time t, there is a centralized market.[1] At $t+1$, there are some frictions that limit investors' ability to trade and contact financial intermediaries. For instance, investors might have the possibility to trade only with dealers in decentralized OTC markets. Or, investors might have no market access at all but the ability to contact an intermediary, with the choice to roll over their deposits or to withdraw them. To this end, we introduce the notation $c_{t+1}\left(\varepsilon, \omega_{t+1}\right)$ where ω denotes the investor's relation with the intermediary, such as whether or not the investor is able to contact an intermediary and trade, or the state-contingent features of the banking contract.

[1] In some of our applications, only a fraction of the investors have access to the market, so the discussion that follows applies to agents who have the ability to trade at time t.

Investors that have access to the time-t centralized market can trade a short-term asset (i.e., liquidity) and a long-term security. The long-term security has a payoff $R_{t+2} > 1$ at $t+2$ and the short-term asset can be consumed at $t+1$ or stored and carried to $t+2$. To simplify the exposition, we assume that the payoff R_{t+2} is deterministic, but the analysis can be extended to the case with a stochastic payoff.

Consider now an investor that is allocating an amount I_t of wealth to liquidity s_t and the long-term asset k_t by trading in the time-t centralized market. The investor is subject to the budget constraint

$$s_t + k_t q_t \leq I_t \tag{4}$$

where q_t is the price of the long-term security.

The investor's optimization implies the stochastic discount factor (SDF) that is used to price, at time t, the long-term security with payoff at $t+2$:

$$SDF = \frac{\partial U / \partial c_{t+2}}{\mathbb{E}_{\varepsilon,\omega}\left\{\partial U / \partial c_{t+1}(\varepsilon,\omega)\right\}}. \tag{5}$$

The denominator of the SDF includes the marginal utility of consumption at time $t+1$ (as opposed to the time-t marginal utility as in standard asset pricing models) because the investor is trading off the investments in the long-term asset—that pays off at time $t+2$—with the investment in liquidity—that can be used for consumption at $t+1$. In addition, the time-$t+1$ marginal utility is weighted with respect to the preference shock ε and, possibly, investors relationship with the intermediary ω, because of the uncertainty associated with these events. We also note that, in the banking application of Section 3, the investor might not always trade in the time-t market—a bank that pools investors' resources and offer demand deposits might be doing the trades. However, as long as the bank acts in the depositors' interest, the bank uses the same SDF in (5).

Using (5), we obtain the asset pricing equation:

$$1 = \frac{{}^{\partial U}\!/\!{}_{\partial c_{t+2}}}{\mathbb{E}_{\varepsilon,\omega}\left\{{}^{\partial U}\!/\!{}_{\partial c_{t+1}(\varepsilon,\omega)}\right\}} \frac{R_{t+2}}{q_t} \tag{6}$$

(where ${}^{R_{t+2}}\!/\!{}_{q_t}$ is the equilibrium gross return on the long-term asset), or, using the linearity assumption in (2), and rearranging:

$$q_t = \beta R_{t+2} \frac{1}{\mathbb{E}_{\varepsilon,\omega}\left\{{}^{\partial U}\!/\!{}_{\partial c_{t+1}(\varepsilon,\omega)}\right\}}. \tag{7}$$

This expression states that the marginal utility of consumption at $t+1$ crucially affects the way the cash-flow R_{t+2} is discounted. First, it is crucial that $U(\cdot)$ is a strictly concave function in c_{t+1}—with linear utility at $t+1$, the price q_t would be just the present-value of the cash-flow, no matter what the distribution of $t+1$ consumption is. Second, the dispersion in marginal utilities at $t+1$ crucially affects discounting. To see this point, consider first the extreme case in which investors' marginal utilities at $t+1$ are equalized for all agents. This is the case if investors are able to trade very easily with intermediaries or are able to withdraw the desired amount of deposits, and it corresponds to the Pareto-optimal allocation described in (3). In this case, the price q_t is given by

$$q_t = \beta R_{t+2}, \tag{8}$$

where we have use the assumption that the Pareto-optimal allocation has unitary marginal utility at $t+1$. In words, the price of the long-term asset is simply the discounted payoff, where the discount rate is given by β. This is the price of the financial asset that prevails in normal times, that is, when intermediaries are well-functioning.

Consider next the case in which investors face difficulties, at time $t+1$ in trading or redeeming their own deposits—because financial intermediaries or financial markets are somewhat impaired.[2] We show next that, in this case, a fire-sale price emerges, that is, the price of the long-term asset is lower than in (8).

The fire-sale price is the result of the malfunctioning financial system, which exposes investors to liquidity risk. More precisely, investors are concerned about facing a liquidity shock ε in states ω in which financial intermediaries are unable to trade or provide liquidity—or more generally, in which the terms of trade or those of the demand-deposit contract are not advantageous. As a result, investors increase their time-t demand for liquidity s_t—to provide self insurance against liquidity shocks, albeit partial—and reduce their demand for the long-term illiquid asset. Provided that the supply of the liquid and long-term illiquid assets are somewhat rigid, these changes in demands reduce the relative price q_t of the illiquid asset, thereby generating a fire-sale price.

To establish the results formally, we impose two sufficient conditions: (i) the economy-wide supply of liquidity s_t and of the long-term asset k_t at time t are fixed (i.e., independent of q_t); and (ii) the malfunctioning in the financial system is extreme, so that no intermediation takes place at $t+1$—investors' only source of consumption at $t+1$ is their own holding of liquidity s_t. Let us stress that these conditions are sufficient, and the results can be derived under somewhat weaker assumptions.

We can then discuss the choices of investors at time t. Investors choose how to allocate their wealth I_t between liquidity s_t and investments in the long-term asset k_t to maximize their utility

$$\max_{s_t, k_t, c_{t+1}(\varepsilon), c_{t+2}} \mathbb{E}_\varepsilon \left\{ U\left(c_{t+1}(\varepsilon), c_{t+2}, \ldots; \varepsilon\right)\right\}$$

subject to the budget constraint (4), the feasibility constraint for time-$t+1$ consumption, $c_{t+1}(\varepsilon) \leq s_t$ for all ε (i.e., they cannot consume more than what they stored at time t), and the feasibility constraint for time-$t+2$ consumption,

$$c_{t+2} \leq R_{t+2} k_t + \left[s_t - c_{t+1}(\varepsilon)\right].$$

[2]These impairments are correctly anticipated as of time t.

Note that the problem is not indexed by ω because we assumed that the malfunctioning in the financial system is extreme (i.e., item (ii) above) so that investors are not able to interact with financial intermediaries at $t+1$ at all. For an investor with a preference shock ε that is sufficiently low, the constraint $c_{t+1}(\varepsilon) \leq s_t$ is not binding, and using (2), the first-order condition is

$$\frac{\partial U\left(c_{t+1}(\varepsilon), c_{t+2}, \ldots; \varepsilon\right)}{\partial c_{t+1}(\varepsilon)} = \beta. \tag{9}$$

This result leads to two observations. First, (9) implies that investors for which the constraint $c_{t+1}(\varepsilon) \leq s_t$ is not binding consume more than in the first-best allocation or more than in normal times—recall that marginal utilities at $t+1$ are normalized to one, whereas the marginal utility in (9) is equal to β, and $\beta < 1$. As a result, fixing total consumption at $t+1$, the other investors—those for which the constraint $c_{t+1}(\varepsilon) \leq s_t$ is binding—must be consuming less than in normal times. Second, because the investors with non-binding $c_{t+1}(\varepsilon) \leq s_t$ constraint are actually storing some of their liquidity holdings, and liquidity holdings are instead entirely used for consumption in normal times, total consumption is actually less than in normal times. Because $\partial U(\cdot)/\partial c_{t+1}$ is convex in c_{t+1}, the two observations lead to the conclusion that

$$\mathbb{E}_{\varepsilon}\left\{\frac{\partial U\left(c_{t+1}(\varepsilon), c_{t+2}, \ldots; \varepsilon\right)}{\partial c_{t+1}(\varepsilon)}\right\} > 1. \tag{10}$$

That is, the increase in marginal utility dispersion together with a reduction in total consumption unambiguously leads to an increase in the average marginal utility in comparison to its normal time value—which is normalized to one.

We can now return to the pricing condition in (7), which implies

$$q_t = \beta R_{t+2} \frac{1}{\mathbb{E}_{\varepsilon,\omega}\left\{{}^{\partial U}\!/\!_{\partial c_{t+1}(\varepsilon,\omega)}\right\}} < \beta R_{t+2},$$

where the inequality follows from (10). Thus, the price of the long-term asset is now less than the normal-time value βR_{t+2} because of the distortion in the pricing kernel that arise from the investors' exposure to liquidity risk.

Sections 3 and 4 present two applications in which the liquidity-risk pricing that generates fire sales is used in the context of panic-based bank runs and a market freeze in OTC trades Before turning to such applications, however, we analyze the efficiency of ex-ante investments in liquid and long-term assets.

2.2 Efficiency of ex-ante investments

We now introduce a $t-1$ choice of investment in the liquid and long-term asset and study whether the investment choice is efficient or if a regulatory intervention can improve welfare. We show that the $t-1$ investment is efficient and, thus, any regulatory intervention that alter such investment decisions backfires, in the sense that it reduces welfare.

At $t-1$, investors have one unit of endowment that can be either invested in liquidity (i.e., stored) or invested in the long-term asset. We denote s_{t-1} to be the investor's choice of liquidity and k_{t-1} the investor's choice of the long-term asset. Because the investor has one unit of endowment, we have $k_{t-1} = 1 - s_{t-1}$, and in what follows, we focus just on the choice of s_{t-1}. The investor solves the problem

$$\max_{0 \leq s_{t-1} \leq 1} \mathbb{E}_{t-1} \left\{ V_t \left(s_{t-1}; q_t \right) \right\} \tag{11}$$

where $V_t \left(s_{t-1}; q_t \right)$ is the indirect utility function of the investor evaluated at time t, when the market price of the long-term asset is q_t. The investor choice of s_{t-1} solve the first-order condition

$$\frac{\partial \mathbb{E}_{t-1} \left\{ V_t \left(s_{t-1}; q_t \right) \right\}}{\partial s_{t-1}} = 0.$$

A regulator that has access to policy tools to alter investors' choice of s_{t-1} maximizes the same objective function as the investor, that is, (11), but with a crucial difference. That is, the regulator internalizes that her choices affect prices—and in particular, the time-t price of the long-term asset, q_t. Hence, the regulator's optimal choice of s_{t-1} solves[3]

$$\frac{\partial \mathbb{E}_{t-1}\left\{V_t\left(s_{t-1};q_t\right)\right\}}{\partial s_{t-1}}+\mathbb{E}_{t-1}\left\{\frac{\partial V_t\left(s_{t-1};q_t\right)}{\partial q_t}\frac{\partial q_t}{\partial s_{t-1}}\right\}=0. \tag{12}$$

To establish that the regulator's choice is the same as the one made by the investor, we need to show that the second term in (12) is zero. To this end, we can write the indirect utility $V_t\left(s_{t-1};q_t\right)$ as

$$V_t\left(s_{t-1};q_t\right)=U\left(\left\{c_{t+1},c_{t+2},...\right\};\varepsilon\right)+\lambda_t\left[I_t-s_t-q_tk_t\right]. \tag{13}$$

The indirect utility depends on two standard element. The first element is the expected utility $\mathbb{E}_{t-1}\left\{U\left(\cdot\right)\right\}$ from the consumption allocation $\left\{c_{t+1},c_{t+2},...\right\}$. The second element is the product of the time-t budget constraint (4)—evaluated with equality—and the relative Lagrange multiplier λ_t. The wealth I_t that the investor allocates to investments in liquidity and long-term assets is given by $I_t=s_{t-1}+q_tk_{t-1}-x_t$, where s_{t-1} and k_{t-1} are the ex-ante investments in liquidity and long-term assets, and x_t denotes other possible uses of liquidity at time t—in one of our application, investors might use x_t to finance some consumption at time t. In the banking application of Section 3, investors might not trade directly and a bank might manage the resources of many investors, but (13) is unchanged because the banks' budget set is obtained by summing over the budget sets of the individual investors that have a relationship with the bank, and because of the linearity of the budget constraints in its entries.

We now have all the elements to show that the second term in (12) is zero. Focusing on the term $\partial V_t\left(s_{t-1};q_t\right)/\partial q_t$, we have

$$\begin{aligned}\mathbb{E}_{t-1}\left\{\frac{\partial V_t\left(s_{t-1};q_t\right)}{\partial q_t}\right\} &= \mathbb{E}_{t-1}\left\{\lambda_t\left[k_{t-1}-k_t\right]\right\}\\ &= 0.\end{aligned}$$

[3]We assume that the indirect utility function is a well-behaved function and we can interchange the derivative with respect to s_{t-1} with the expectation operator.

The second line follows from the time-t market clearing condition for capital $k_t = k_{t-1}$—the assumption introduced in Section 2.1 about the fixed supply of the assets traded at time t implies that the supply of the long-term asset is determined at time $t-1$.[4]

When our framework is compared with other models in the fire sale literature, a key difference is that the price of the long-term asset q_t does not enter any borrowing constraint. As a result, the pecuniary externalities that are at work when such constraint is present (Dávila and Korinek 2017) are absent here. We summarize this result in the next proposition.

Proposition 2.1 *The choice of liquidity s_{t-1} and investments in the long-term assets $k_{t-1} = 1-s_{t-1}$ made by price-taking agents at time $t-1$ is the same as the choice made by a regulator that internalizes the effects on prices.*

The result of this section has important policy implications. The ex-ante investment decisions of price-taking agents that face possible fire-sale prices are efficient, and regulation aimed at changing such decisions is counterproductive, in the sense that reduces welfare. For instance, in the context of the banking application of Section 3, the result of Proposition 2.1 implies that regulatory interventions in the form of liquidity requirements are not needed, and if implemented, are welfare reducing, when fire sales and fire-sale prices are driven by the liquidity-risk pricing mechanism studied in this paper.

3 First Application: Bank Runs

We now present our first application. We extend the framework of Diamond and Dybvig (1983) (DD) to endogenize the fire-sale price at which banks sell their long-term assets in the event of a run according to the mechanism proposed in Section 2. We try to keep the model as close as possible to DD to highlight the similarities and differences and to simplify the exposition by building on what has become a standard framework in the literature. We

[4] It is straightforward to extend the framework to introduce a technology that allows to liquidate the capital at t. As long as there are liquidation costs, our results are unchanged.

first present the results in a framework in which the demand deposit banking contract is imposed exogenously, and then provide extensions in which we endogenize some elements of the contract.

3.1 Environment

The economy lasts four periods denoted by $t \in \{0, 1, 2, 3\}$. As the rest of the discussion will clarify, the structure and assumption of time $t = 0$, $t = 2$, and $t = 3$ are very similar to the original DD paper, whereas time $t = 1$ represents a departure that we use to model our novel fire-sale mechanism described in Section 2.

There is a continuum of agents that are ex-ante identical (i.e., identical at $t = 0$) and are subject to preference shocks. At $t = 1$, some of the uncertainty related to the preference shocks is realized, and the ex-ante identical agents are split into two groups: a fraction θ of them become *impatient* and the remaining $1 - \theta$ become *non-impatient*. Impatient agents enjoy only utility from consumption at $t = 1$ according to $u\left(c_1\right)$. Non-impatient agents face additional preference shocks at $t = 1$, as in Diamond and Dybvig (1983). In particular, a non-impatient agent becomes *normal* with probability $\frac{\gamma}{1-\theta}$ and *patient* with probability $\frac{1-\gamma-\theta}{1-\theta}$, with utility function

$$u\left(c_1, c_2; type\right) = \begin{cases} u(c_1) + \beta c_2 & \text{if } type = \{\text{normal}\} \\ \beta c_2 & \text{if } type = \{\text{patient}\}. \end{cases}$$

We assume throughout the analysis that

$$u\left(c\right) = \log c. \tag{14}$$

While the formulation of preferences is slightly different from Diamond and Dybvig (1983), the logic and implications are very similar. That is, impatient and normal agents have urgent needs to consume at $t = 1$ and $t = 2$, respectively.

All the preference shocks are i.i.d. across households, and the law of large numbers holds. As a result, the overall mass of impatient, normal, and patient agents is θ, γ, and $(1-\gamma-\theta)$, respectively. We assume throughout the analysis that

$$u(c) = \log c. \tag{15}$$

At $t = 0$, each agent is endowed with 1 unit of the economy's single good at time 0. These goods can either be stored or invested in a long-term technology—each unit of investment is referred to as "capital." The storage technology is standard and allows agents to transfer goods from time t to $t+1$ with a gross return of one. The long-term technology allows agents to invest at $t = 0$, in projects that produce outputs at $t = 3$. Each unit of the endowment invested produces $R > 1$ units of output at $t = 3$, with the normalization $R = 1/\beta$. We depart from Diamond and Dybvig (1983) by assuming that there is no liquidation technology to transform the long-term investments into consumption goods after the time-0 decision has been made.

A key element of our environment is that agents have access to a centralized location at $t = 1$ where they can trade capital against stored goods. More precisely, we assume that a fraction p of agents have access to the time-1 centralized market, and the remaining fraction $1-p$ do not. We denote q_1 to be the price of capital in this market. We also assume that there is no market at $t = 2$. The assumptions of limited market participation at $t = 1$ and absence of centralized markets at $t = 2$ prevent agents from using trading in financial markets as a way to obtain liquidity in the event of preference shocks, and opens up a role for banks (Jacklin, 1987).

Banks can be formed at $t = 0$ and liquidated at $t = 3$. After banks are formed at $t = 0$, each agent can contact her bank at all $t \in \{1, 2, 3\}$. In particular, in keeping with the assumption that there is no centralized market at $t = 2$, we assume that only bilateral meetings between each depositor and her bank are possible in that time period, so that depositors cannot meet with each other but only with their own bank at $t = 2$.

3.2 Pareto optimal allocation

We first present the Pareto optimal allocation in the absence of any information friction. That is, we consider a social planner who can observe individuals' realized types at $t = 1$ and $t = 2$. At $t = 0$, the planner stores s_0 and invests $1 - s_0$ in the long-term technology. The planner let each impatient agent consume c_1 at $t = 1$, each normal agent consume c_2 at $t = 2$, and each patient agent consume c_3 at $t = 3$. Hence, the planner solves

$$\max_{s_0,c_1,c_2,c_3} \; \theta u(c_1) + \left[\frac{\gamma}{1-\theta}\, u(c_2) + \frac{1-\gamma-\theta}{1-\theta}\beta c_3\right]$$

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

$$\theta c_1 + (1-\theta)\,\gamma c_2 \leq s_0,$$

$$(1-\theta)\,(1-\gamma)\,c_3 \leq (1-s_0)R.$$

The optimal consumption c_1^* and c_2^* satisfy

$$u'(c_1^*) = u'(c_2^*) = 1$$

and the assumption of log utility in (14) implies

$$s_0^* = \theta + \gamma, \quad c_1^* = 1, \quad c_2^* = 1, \quad c_3^* = R.$$

The next proposition summarizes this result.

Proposition 3.1 (Pareto optimal allocation) *The Pareto optimal allocation features* $s_0^* = \theta + \gamma$, $c_1^* = c_2^* = 1$, *and* $c_3^* = R$.

3.3 Decentralized economy with restricted banking contracts

We study a decentralized equilibrium, beginning with the analysis of a banking contract that implements the Pareto optimal allocation of Section 3.2. This is a standard "good equilibrium" of banking model in the tradition of Diamond and Dybvig (1983). That is, the bank collects endowments from agents and invest $1 - s_0$ in the long-term technology, and offers a deposit contract that provides full insurance against preference shocks. We then turn to the analysis of the bad equilibrium in Section 3.3.2.

In this section, we focus on the case in which all agents have market access at $t = 1$ (i.e., $p = 1$) and restrict the set of contracts that banks can offer to provide a simple exposition that focuses on our novel contribution. In particular, We assume that banks are set up at $t = 0$ and offer a contract that allow agents to withdraw at $t = 1$ or $t = 2$ or $t = 3$. In other words, we ignore that the preference shocks realized at $t = 1$ can be self-insured by accessing the time-1 centralized market (Jacklin, 1987). We extend the analysis in Section 3.4, in which we consider $p < 1$ (i.e., limited market participation) and the possibility to set up banks not only at $t = 0$ but also at $t = 1$, showing that the main results are unchanged.

As an additional restriction on contracts, we assume that tools such as suspension of convertibility or deposit insurance are not available. This is again made to better convey our intuition, and builds on well-established results in the literature that highlight how these tools are not necessarily optimal if e.g. banks cannot commit to suspend convertibility (Ennis and Keister, 2009) or if deposit insurance leads to moral hazard (Cooper and Ross, 1998).

An equilibrium with banks is defined as a contract that specifies banks' investments in storage $\hat{s}_0$ and in the long-term technology $1 - \hat{s}_0$, and withdrawals $\hat{c}_1$, $\hat{c}_2$, and $\hat{c}_3$ for agents that report their type to be impatient, normal, or patient, respectively; agents' and banks' decisions about the quantity of storage and long-term assets that each of them want to trade at $t = 1$; and a price q_1 that clears the market at $t = 1$.

3.3.1 Good equilibrium

The good equilibrium is standard. Banks offer a contract in which they collect endowments at $t = 0$ and allow agents to withdraw at $t = 1$, or $t = 2$, or $t = 3$. In equilibrium, impatient agents withdraw at $t = 1$, normal agents withdraw at $t = 2$, and patient ones wait until $t = 3$ and receive an equal share of the funds that are available at that time. The Pareto optimal allocation presented in Section 3.2 is implemented, as summarized by the next proposition.

Proposition 3.2 (Good equilibrium) *There exists an equilibrium in which banks offer a contract that implements the Pareto optimal allocation of Prop 3.1, that is,* $\hat{s}_0 = \theta + \gamma$, $\hat{c}_1 = \hat{c}_2 = 1$, *and* $\hat{c}_3 = R$*; no trading takes place at* $t = 1$*; and the price of capital is* $q_1 = 1$.

The proof (in Appendix A) is based on two steps. The first one is the standard argument that depositors prefer to truthfully report the realization of their preference shock. The other steps deal with trading decisions and the time-1 market, which is novel in our environment, so we elaborate more on it.

To show that the time-1 market clears at the price $q_1 = 1$, we need to solve for banks' and agents' trading decisions. For impatient agents, consuming all their withdrawal leads to a marginal utility of one at $t = 1$, which is equal to the marginal utility of investing a dollar in the long-term asset and using the proceeds for consumption at $t = 3$. Non-impatient agents do not withdraw at $t = 1$ and, thus, have no resources to trade. We then need to show that banks do not want to adjust the composition of their holdings of storage and illiquid capital at $t = 1$. After withdrawals by impatient agents took place, banks' have the option to adjust their holdings of liquidity, s_1, and capital, k_1, by engaging in trades. Their objective is to maximize the utility of non-impatient agents:

$$\max_{s_1, k_1, \hat{c}_2, \hat{c}_3} \frac{\gamma}{1-\theta} u\left(\hat{c}_2\right) + \frac{1-\gamma-\theta}{1-\theta} \beta \hat{c}_3$$

subject to

$$s_1 + q_1 k_1 \leq s_0 - \theta + q_1\left(1 - s_0\right) \tag{16}$$

$$\gamma \hat{c}_2 \leq s_1$$

$$(1 - \gamma - \theta)\, \hat{c}_3 \leq R\, k_1.$$

That is, a bank trades with the objective of improving the terms of the contract, subject to the budget constraint at $t = 1$, (16), the constraint that time-2 withdrawals $\hat{c}_2$ to the mass γ of normal agents must be paid using the goods s_1 stored at $t = 1$, and the constraint that time-3 consumption to patient agents is paid using the output produced by the capital k_1 that the bank has after trading at $t = 1$.

Under the equilibrium price $q = 1$, banks' maximization problem implies

$$s_1 = \gamma, \quad k_1 = 1 - \hat{s}_0.$$

That is, banks optimal holdings of capital, k_1, do not change in comparison to the time-0 holdings $1 - \hat{s}_0$. In other words, banks do not want to engage in any trade at $t = 1$.

To further clarify the result, it is useful to restate the optimality condition of banks' trading decisions as

$$1 = \beta \left[\frac{v'(\hat{c}_2)}{u'(\hat{c}_1)} (1 + r) \right] \tag{17}$$

where $v(c) = c$ is agents' utility at $t = 2$ and $1 + r \equiv \frac{R}{q}$ is the return on purchasing a unit of capital. This is the same asset pricing condition derived in Section 2, and when evaluated at the equilibrium level of consumption, it implies

$$q_1 = \beta R. \tag{18}$$

Because of the normalization $R = 1/\beta$, the result $q_1 = 1$ follows.

3.3.2 Bad equilibrium

As in the DD paper, there can be another equilibrium in which bank runs become self-fulfilling prophecies. In this equilibrium, banks are subject to a run at $t = 1$ and sell their holdings of the long-term assets to repay the withdrawals. Crucially, the long-term asset is sold by banks at a fire-sale price $q < 1$ (i.e., lower than the good equilibrium price). The fire sale price arises here because, as banks are subject to runs and fail at $t = 1$, they are unable to provide insurance against the liquidity risk at $t = 2$. Hence, agents' increased demand for liquidity and decreased demand for the long-term illiquid asset at $t = 1$ put downward pressure on the price q_1. This lower price, in turn, implies that banks are insolvent at $t = 1$, making runs a self-fulfilling outcome.

In this equilibrium, the bank still collects endowments from agents and invests $1 - s_0$ in the long-term technology and provides the same deposit contract. At $t = 1$, all agents go to the bank and withdraw. The amount each agent is able withdraw from the bank is

$$w(q_1) = s_0 + (1 - s_0)q_1.$$

Agents adjust their portfolio at the same time when they withdraw. For an impatient type, let c_i be the amount she consumes out of $w(q)$, which solves

$$\max_{c_i} \ \log(c_i) + \beta \frac{w(q_1) - c_i}{q_1} R.$$

The FOC gives

$$c_i^* = q_1.$$

If $q_1 < 1$, we have $c_i^* < 1 = c_1^*$. Accordingly, an impatient type consumes less than she does in the efficient equilibrium and purchases capital if she observe a lower price. For a non-impatient agent, let s_n be the amount of storage goods she carries (which she does not consume immediately), which solves

$$\max_{s_n} \ (1 - \mu) \left[\log(s_n^*) + \beta \frac{(w(q_1) - s_n)R}{q_1} \right] + \mu\beta \left[s_n + \frac{(w(q) - s_n)R}{q_1} \right]$$

where $\mu \equiv \frac{1-\theta-\gamma}{1-\theta}$. With probability $1-\mu$, the non-impatient agent becomes normal at time 1; with probability μ she turns out being patient. The storage s_n is determined by the following FOC

$$\frac{1-\mu}{s_n^*} = \frac{1}{q_1} - \mu\beta,$$

or

$$s_n^* = \frac{1-\mu}{\frac{1}{q_1} - \beta\mu}.$$

In equilibrium, the storage goods market and the capital market must clear. By Walras' Law, we only need to check the storage goods market. The clearing condition for this market writes as

$$\theta c_i^* + (1-\theta) s_n^* = s_0 \tag{19}$$

which determines the market price q_1. We now show that $q_1 < 1$ in this equilibrium.

Proposition 3.3 *In the bank run equilibrium, there is fire sale:* $q_1 < 1$*.*

Proof: The market clearing condition (19) simplifies as

$$\theta q_1 + \frac{\gamma}{\frac{1}{q_1} - \beta\mu} = \theta + \gamma$$

Suppose $q_1 \geq 1$, then $\theta q_1 + \frac{\gamma}{\frac{1}{q_1}-\beta\mu} > \theta + \gamma$, the market cannot clear, a contradiction. □

3.4 Extension: limited market participation

The previous section has shown the existence of a bad equilibrium in which the fire-sale price at which banks sell their investments is endogenously determined according to a standard consumption-based asset pricing formulation. However, the results are derived in a setting in which agents can gather in a centralized market at time $t = 1$ in which their only option is to trade. In practice, however, when a run takes place, some depositors who run might move their resources to other parts of the financial sector that allows them to receive full or almost-full insurance against their liquidity risk. This possibility would affect depositors' pricing kernel and limit the fire-sale mechanism that we have analyzed. Nonetheless, we show that as long as some agents face frictions in their ability to allocate their financial investments, the bad equilibrium with fire sale exists. In particular, we assume that *some* depositors face trading frictions or other costs that limit their ability to trade and to move their resources to other financial intermediaries.

We capture this idea by assuming a form of limited participation in centralized market. That is, only a fraction $p \leq 1$ of agents can access the market, and the remaining $1-p$ have only access to the bank established at $t = 0$.

Appendix B analyzes the case in which all agents can participate in the market at $t = 1$ (i.e., $p = 1$) and pool their resources to create new banks. In that case, the bad equilibrium does not exist. Here, we analyze the case in which $p < 1$ and, thus, some agents can only, at $t = 1$ withdraw from their pre-existing bank and store liquidity. In this case, a bad equilibrium with a fire-sale price still exists.

Assume that agents enter the time-1 centralized market (hence are able to trade or create a bank) with probability $p \leq 1$. With probability $1-p$, agents are in an autarky situation. Note that in any case, all agents would be able to withdraw from the bank that was set up at $t = 0$.

The introduction of the additional layer of decentralization does not change the Pareto optimal allocations. In the good equilibrium in which the Pareto optimal allocations are implemented, if a non-impatient type deviates and withdraws $c_1^* = 1$ at time 1. With probability p, she is able to adjust her stock of storage goods to $s_n^* = \frac{1-\mu}{1-\beta\mu}$. With probability $1-p$, she is isolated from trading. In this autarky case, she does not consume any goods at time 1 but carries $c_1^* = 1$ to time 2. If, at time 2, she becomes a normal type, she chooses c_n^a

to consume at $t = 2$, and $1 - c_n^a$ to store. Her maximization problem at $t = 2$ is

$$\max_{c_i^a} \ \log(c_n^a) + \beta(1 - c_n^a)$$

and she chooses $c_n^a = 1$ at the corner because $1 < \frac{1}{\beta}$. If she becomes patient at time 2, she will just consume 1 at time 3. So the expected utility of an isolated non-impatient type at $t = 1$ is

$$(1 - \mu)\log(c_1^*) + \beta\mu c_1^*$$

The incentive constraint for non-impatient types to not withdraw at time 1 becomes

$$\begin{aligned}(1-\mu)\log(c_2^*) + \mu\beta R \geq p\Big((1-\mu)\log(s_n^*) + \beta\big(\mu s_n^* + (c_1^* - s_n^*)R\big)\Big) \\ +(1-p)\Big((1-\mu)\log(c_1^*) + \beta\mu c_1^*\Big) \qquad (20)\end{aligned}$$

The expected utility of a non-impatient type under autarky is lower than under free trading, hence a non-impatient agent will have lower incentive to deviate when she knows that she may not be able to reallocate any capital. We have

Proposition 3.4 *The good equilibrium always exists with limits on market participation.*

Proof: The old incentive constraint (27) holds implies that (20) holds. The result follows. □

The bad equilibrium with bank run differs from the one in the previous discussion. In autarky, an agent who withdraws $w(q)$ from the old bank can only adjust intertemporal consumption. Namely, an impatient type at $t = 1$ chooses c_i^a to consume at $t = 1$ and stores the rest $w(q) - c_i^a$ until $t = 3$. The consumption c_i^a solves

$$\max_{c_i^a} \ \log(c_i^a) + \beta(w(q) - c_i^a)$$

which has the following formula

$$c_i^a = \begin{cases} \frac{1}{\beta} & \text{if } \frac{1}{\beta} \leq w(q) \\ w(q) & \text{if } \frac{1}{\beta} \geq w(q) \end{cases}$$

When $w(q)$ is greater than $\frac{1}{\beta}$, the FOC implies $c_i^a = \frac{1}{\beta}$. When $w(q)$ is below $\frac{1}{\beta}$, the solution is at the corner. Similarly, at $t = 1$, a normal type chooses c_n^a to consume at $t = 2$, and $w(q) - c_n^a$ to store. The solution of c_n^a is identical to c_i^a

$$c_n^a = \begin{cases} \frac{1}{\beta} & \text{if } \frac{1}{\beta} \leq w(q) \\ w(q) & \text{if } \frac{1}{\beta} \geq w(q) \end{cases}$$

A patient type at $t = 2$ will consume zero at time 2 and $w(q)$ at time 3. Here when $q \leq 1$, $w(q) \leq 1$, and $c_i^a = c_n^a = w(q)$.

Non-impatient agents that successfully access to the centralized market create a new bank that provides insurance against liquidity risk at $t = 2$. The bank, collecting $p(1 - \theta)w(q)$ from those non-impatient agents will stick to a investment plan in storage goods

$$s_1 = \frac{\gamma}{p(1 - \theta)w(q)}$$

The deposit contract remains the same, with $c_2' = 1$ and $c_3' = R$.

In sum, at $t = 1$, an impatient-type agent: with probability p, she consumes q and invests $\frac{w(q)-q}{q}$ in capital; and with probability $1 - p$, she withdraws $w(q)$ and consumes c_i^a at $t = 1$ and the rest at $t = 3$. And a non-impatient type: with probability p, she is able to create a bank with other non-impatient agents, which invests $s_1 p(1 - \theta)w(q)$ in storage goods; with probability $1-p$, she withdraws $w(q)$ and consumes c_n^a at $t = 2$ if she is normal and consumes $w(q)$ at $t = 3$ if she is patient.

The demand for liquidity of an autarky agent at $t = 1$ is just $w(q)$. By LLN, the market clearing condition for storage goods now writes as

$$p\theta q + p\gamma + (1-p)(s_0 + (1-s_0)q) = \theta + \gamma \tag{21}$$

The first term in (21) is the aggregate demand from impatient types who are able to adjust portfolios. The second term is the demand from the new bank created, which is just γ. The third term is the total amount of withdrawals by autarky agents. The equilibrium capital price is

$$q = \frac{p\theta}{p\theta + (1-p)(1-\theta-\gamma)}$$

As long as $p < 1$, q is smaller than one, implying there is fire sale.

Proposition 3.5 *In the bank run equilibrium, there is fire sale: $q < 1$.*

4 Second Application: Trading in OTC Markets

As a second application, we consider a very simple stylized model of OTC trades. A fire sale here is generated by an exogenous shock that changes the fundamentals, as opposed to a panic as in the previous example. In this sense, this second application highlights the fact that our theory of fire-sale price applies not just to financial crises driven by panics but also to disruptions in the financial sectors that are caused by so-called fundamental shocks.

For this application, we employ a simple three-period model $t = 1, 2, 3$. The economy is populated by investors endowed, at $t = 1$, with liquidity s——that can be stored and used for consumption at $t = 2$ or $t = 3$—and a long-term illiquid asset k—that produces a return R in the last period, that is, $t = 3$. Investors have preferences

$$\varepsilon u(c_2) + \beta c_3, \tag{22}$$

where c_2 and c_3 are consumption at $t = 2$ and $t = 3$, respectively, and ε is a preference shock that capture the notion that investors have different liquidity needs—see e.g. the discussion in Lagos, Rocheteau, and Weill (2011). To keep the analysis simple, we assume that ε can take two values: $\varepsilon = 0$ with probability $1 - \theta$ and $\varepsilon = 1$ with probability θ. We impose the normalization that the endowment of storage is $s = \theta$ per investor, and that $R = 1/\beta$. We also assume that $u(\cdot) = \log(\cdot)$.

At $t = 1$, there is a centralized market in which the asset k can be traded at price q_1. At $t = 2$, there is no centralized market, and investors can only trade with dealers in an OTC market. We assume that each investor is able to contact one dealer with probability $\alpha \in [0, 1]$. Dealers can trade with each other in an interdealer market at price q_2, similar to Duffie, Gârleanu, and Pedersen (2005). Dealers have no utility from holding the asset, but they only have linear utility c_2^d from consuming liquidity. We make the assumption that, in a meeting between an investor and the dealer, the investor gets all the surplus (i.e., investors make a take-it-or-leave-it offer).

4.1 Equilibrium with a well-functioning OTC market

We now characterize the equilibrium—and in particular, the price q_1 of the long-term asset—when dealers are well-functioning. We consider the limiting case in which $\alpha = 1$, so that investors are able to meet a dealer with certainty.

Because investors extract all the surplus from the meeting, it is as if investors were able to trade the long-term asset in the interdealer market at price q_2. We thus solve the problem, for an investor that enter time $t = 2$ with holdings k_1 and s_1:

$$\max_{k_2 \geq 0, c_2 \geq 0, c_3 \geq 0} \varepsilon u\left(c_2\right) + c_3 \tag{23}$$

subject to the budget constraint $c_2 + q_2 k_2 \leq q_2 k_1 + s_1$ and to $c_3 = R k_2$.

For an investor with preference shock $\varepsilon = 1$, we have

$$u'(c_2) = \beta \frac{R}{q_2}. \tag{24}$$

In equilibrium, all the storage is consumed by investors with realized preference shock $\varepsilon = 1$—the other investors have no utility from consumption at $t = 1$. Hence, (24) and the normalization that the overall supply of liquidity is $s = \theta$ and that $R\beta = 1$ imply $q_2 = 1$.

Stepping back to $t = 1$, we observe that the price of the long-term satisfy $q_1 = q_2$ and, thus, $q_1 = 1$. This is because no matter what the realization of the preference shock is, investors at $t = 2$ have unitary marginal utility from wealth.

Note that the result is quite similar to our bank run application. Indeed, the well-functioning market at $t = 2$ essentially provides investors with full insurance against liquidity risk—a well-known result derived by Jacklin (1987).

4.2 Equilibrium with disruptions in the OTC market

Next, we consider the other extreme case in which the OTC market is plagued by extreme frictions. In particular, we study the case $\alpha = 0$, that is, investors are unable to meet dealers to trade at $t = 2$.

The framework maps directly in the bad equilibrium analyzed for the banking application in Section 3.3.2—with a small distinction driven by the fact that there are no impatient consumers here that carry some liquidity to $t = 3$. Indeed, both cases are characterized by the absence of markets at $t = 2$.

At $t = 1$, investors choose the optimal holdings of liquidity s_1 and long-term asset k_1 to solve

$$\max_{k_1, s_1} (1 - \theta) \beta [s_1 + Rk_1] + \theta [\varepsilon u(s_1) + \beta Rk_1] \tag{25}$$

subject to $s_1 + q_1 k_1 \leq s + q_1 k$. The first-order condition, together with the market clearing

$s_1 = \theta$, imply

$$q_1 = \frac{1}{(1-\theta)\,\beta + \theta \times (1/\theta)} < 1. \tag{26}$$

Equation (26) implies that the price of the long-term asset at $t = 1$ is less than one, that is, a fire-sale price arises.

5 Conclusion

This paper provides a novel theory of fire sales in which liquidity risk both justifies the existence of financial intermediaries and generates a fire-sale price when the financial sector is in distress. The approach is quite general and can be applied to several settings, such as traditional models of bank runs as well as models of OTC trade. Crucially, the policy implications differ substantially from those derived in models that use traditional approaches to generate fire sales such as cash-in-the-market pricing, second-best use, and asymmetric information. While ex-ante investments are typically inefficient in those settings, they are efficient here, in the sense that a regulator does not want to impose liquidity requirements because of the costs associated with reducing investments in assets that are more productive—albeit illiquid.

References

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

Cochrane, John H. 2011. "Presidential address: Discount rates." *The Journal of finance* 66 (4):1047–1108.

Cooper, Russell and Thomas W Ross. 1998. "Bank runs: Liquidity costs and investment distortions." *Journal of monetary Economics* 41 (1):27–38.

Dávila, Eduardo and Anton Korinek. 2017. "Pecuniary externalities in economies with financial frictions." *The Review of Economic Studies* 85 (1):352–395.

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

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen. 2005. "Over-the-counter markets." *Econometrica* 73 (6):1815–1847.

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

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

Kiyotaki, Nobuhiro and John Moore. 1997. "Credit cycles." *Journal of political economy* 105 (2):211–248.

Kurlat, Pablo. 2016. "Asset markets with heterogeneous information." *Econometrica* 84 (1):33–85.

———. 2021. "Investment externalities in models of fire sales." *Journal of Monetary Economics* 122:102–118.

Lagos, Ricardo, Guillaume Rocheteau, and Pierre-Olivier Weill. 2011. "Crises and liquidity in over-the-counter markets." *Journal of Economic Theory* 146 (6):2169–2205.

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

Lorenzoni, Guido. 2008. "Inefficient credit booms." *The Review of Economic Studies* 75 (3):809–833.

Merrill, Craig B, Taylor D Nadauld, René M Stulz, and Shane M Sherlun. 2021. "Were there fire sales in the RMBS market?" *Journal of Monetary Economics* 122:17–37.

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

Shleifer, Andrei and Robert W Vishny. 1992. "Liquidation values and debt capacity: A market equilibrium approach." *The Journal of Finance* 47 (4):1343–1366.

Weill, Pierre-Olivier. 2020. "The search theory of over-the-counter markets." *Annual Review of Economics* 12:747–773.

A Appendix: Proofs

Proposition A.1 *The good equilibrium (with restricted banking contract) always exists.*

Proof: We first claim that the price of capital at $t = 1$ has to be $q = 1$.

Lemma A.1 *The equilibrium capital market price is $q = 1$.*

Proof: At $q = 1$, impatient agents will withdraw and consume c_1^* at $t = 1$ without purchasing any capital. To see this, let c_i be the amount of goods an impatient type consumes, hence she purchases $c_1^* - c_i$ as capital that will deliver R in the future. Then, c_i solves the expected utility of an impatient agent

$$\max_{c_i} \ u(c_i) + \beta(c_1^* - c_i)R$$

The FOC derives as $u'(c_i) = 1$, implying $c_i^* = c_1^* = 1$. □

In equilibrium, the incentive compatibility of non-impatient types has to be satisfied. At $q = 1$, suppose a non-impatient type deviates and withdraws c_1^* at time $t = 1$. She can use a portion of the goods to purchase capital in the market and carry the rest to the future. Let s_n be the amount she stores, and $c_1^* - s_n$ be the amount of capital she purchases at the price of 1. At time 2, if she becomes a normal type, she consumes s_n and will consume $(c_1^* - s_n)R$ at time 3; if she becomes patient, she will consume $s_n + (c_1^* - s_n)R$ at time 3. Hence, at $t = 1$, the non-impatient agent is choosing s_n to solve

$$\max_{s_n} \ \frac{\gamma}{1-\theta}\big[\log(s_n) + \beta(c_1^* - s_n)R\big] + \frac{1-\theta-\gamma}{1-\theta}\beta\big[s_n + (c_1^* - s_n)R\big]$$

where $\frac{\gamma}{1-\theta}$ is the conditional probability of being a normal type at time 2 conditioning on being non-impatient at time 1. Let $\mu \equiv \frac{1-\theta-\gamma}{1-\theta}$ as in the main text, the optimal s_n^* is determined by the following FOC

$$(1-\mu)\left(\frac{1}{s_n^*}-1\right)=1-\beta,$$

or $s_n^*=\frac{1-\mu}{1-\beta\mu}$. If the non-impatient type follows the equilibrium path, she will receive and consume c_2^* at $t=2$ if she turns out to be of normal type and consume $\frac{(1-s_0)R}{1-\theta-\gamma}$ at $t=3$ if she becomes patient. The final realization of the investment technology, $(1-s_0)R$, will be evenly distributed among the patient agents. Since $s_0=\theta+\gamma$, each patient agent receives R. The incentive constraint hence writes as

$$(1-\mu)\log(c_1^*)+\mu\beta R\geq(1-\mu)\log(s_n^*)+\beta\left[\mu s_n^*+(c_{\frac{1}{2}}^*-s_n^*)R\right] \tag{27}$$

which becomes

$$0\geq(1-\mu)\log\left(\frac{1-\mu}{1-\beta\mu}\right)$$

after simplification. This condition always holds because $\beta<1$.

Moreover, the equilibrium allocations must guarantee that patient type incentive compatibility is met at time 2. If a patient agent goes to the bank at $t=2$ to withdraw early, she receives c_2^* amount of goods and stores them until time 3 because there is no market at $t=2$. By deviation, she is unable to receive the investment payoff she could have obtained at $t=3$, which amounts to R. The denominator is the fraction of agents who have not withdrawn at time 3. The IC thus writes as

$$c_2^*\leq R. \tag{28}$$

Since $c_2^*=1$, (28) always holds as well.

B Appendix: Full market participation and new banks at $t = 1$

This appendix shows that, when all agents can access the market at $t = 1$ and pool their resources to create new banks, there are no fire sales and the allocation is the same as in the good equilibrium.

The new bank established by the $1 - \theta$ non-impatient agents will collect $(1 - \theta)w(q)$ in total amount of the goods. The bank adjusts its portfolio at the same time. Let s_1 be the fraction the bank carries in storage goods and $1 - s_1$ be the fraction of purchasing capital. The new bank also provides a deposit contract that allows agents to withdraw c_2' at time 2 and c_3' at time 3 if the agent has not withdrawn yet. The new bank is solving

$$\max_{s_1} \ \gamma \log(c_2') + \beta(1 - \theta - \gamma)c_3'$$

subject to

$$\gamma c_2' = s_1(1 - \theta)w(q)$$
$$(1 - \theta - \gamma)c_3' = (1 - s_1)(1 - \theta)w(q)R$$

The investment plan of the bank is

$$s_1 = \frac{\gamma}{(1 - \theta)w(q)}$$

Impatient type agents each withdraws c_1^* from the initial bank and consumes $c_i^* = q$, as shown in the main part of the paper. The market clearing condition for the storage goods market is hence

$$\theta q + (1 - \theta)s_1 w(q) = s_0 \tag{29}$$

implying that the market price has to be $q = 1$. In this case, $c_1^* = c_2' = 1, c_3' = R$. The ex ante expected payoff of an agent is the same as in the good equilibrium. In other words, the two equilibria are isomorphic in regard with welfare.