What Quantity of Reserves Is Sufficient? *

Yilin (David) Yang†

May 9, 2024

Abstract

What quantity of reserves should the Fed supply to support effective monetary policy implementation and an efficient interbank payment system? To answer this question, I construct a model linking interbank intraday payment timing with monetary policy implementation. I show that a low reserve supply causes banks to delay payments to each other and strategically hoard reserves, which in turn disincentivizes banks from providing liquidity to short-term funding markets, driving up the spreads between overnight risk-free market rates and the central bank deposit rate, impeding monetary policy implementation. As reserve balances get sufficiently low, even small reductions in reserves can have large impacts on these spreads, mirroring the events observed in September 2019. Fitted to data from 2019, my model predicts the funding rate spikes of September 16-18, 2019 as an out-of-sample event. The model also provides a counterfactual analysis of the sufficient reserve level that could have prevented the September 2019 repo spike, offering insights into the current discussions about the appropriate size of the Federal Reserve's balance sheet and quantitative tightening (QT).

Keywords: Quantitative Tightening, Monetary Policy Implementation, Financial Stability, Interbank Payment, Strategic Complementarity.

---

*First Draft: August 20, 2020. The most recent version is available here. A previous version of this paper was circulated under the title "Why Repo Spikes" (circulated since 2020). I am grateful to Darrell Duffie for his invaluable advising. I also thank Gara Afonso, Anirudha Balasubramanian, Adam Copeland, Daniel Chen, Adrien d'Avernas (discussant), Sebastian Di Tella, Wenxin Du, Benjamin Hébert, Jonathan Wallen, Arvind Krishnamurthy, Robert Jay Kahn, Xu Lu, Angela Maddaloni (discussant), Marco Machiavelli (discussant), Suraj Malladi, Antoine Martin, Jamie McAndrews (discussant), Monika Piazzesi, Claudia Robles-Garcia, Lorenzo Rigon, Amit Seru, Quentin Vandeweyer, and Jeffrey Zwiebel for helpful discussions. In addition, I thank participants of the finance seminar at Johns Hopkins Carey Business School, 2022 WFA Meeting, the 7th Joint Bank of Canada and Payments Canada Symposium, the mini symposium on funding market liquidity at Stanford University and the OFR Ph.D. Symposium on Financial Stability. All the remaining errors are my own.

†City University of Hong Kong. 83 Tat Chee Ave, Hong Kong. Email: yiliyang@cityu.edu.hk

Our goal is to provide an ample supply of reserves to ensure that control of the federal funds rate and other short-term interest rates is exercised primarily by setting our administered rates and not through frequent market interventions [...] it is clear that without a sufficient quantity of reserves in the banking system, even routine increases in funding pressures can lead to outsized movements in money market interest rates. This volatility can impede the effective implementation of monetary policy, and we are addressing it.

—Jerome Powell, "Data-Dependent Monetary Policy in an Evolving Economy," October 08, 2019[1]

1 Introduction

What quantity of central bank deposits ("reserves") is sufficient to support effective monetary policy implementation and an efficient interbank payment system? As central banks, particularly the US Federal Reserve, continue with balance sheet runoffs, the imperative to understand this question has become ever more critical. This paper sheds light on this issue through a quantitative framework based on game-theoretic modeling of the strategic interactions of large US dealer banks. A low supply of reserves causes banks to delay payments to each other, reinforced in equilibrium by strategic complementarity in interbank intraday payment timing, leading to strategic hoarding of reserves. This in turn disincentivizes banks from providing liquidity in wholesale funding markets, driving up the spreads between wholesale funding market rates and the central bank's deposit rate, referred to as "IOR,"[2] thereby disrupting the efficacy of monetary policy implementation under the prevailing floor system. My model demonstrates that when reserve balances fall below a critical threshold, even minor reductions can substantially impact these spreads, a dynamic that cannot be captured by linear regression models. Using data from 2019, I employ a method of moments procedure to quantify the model. My model successfully predicts, out-of-sample, a liquidity crunch in the wholesale funding markets beginning on September 16, a prediction corroborated by the famous repo-spike event of that year (Fig. 6). Additionally, the model provides an estimate of the minimum level of reserves that hypothetically could have prevented the September 2019 repo spike (Fig. 8)—a level that would support effective monetary policy implementation, maintain an efficient interbank payment system, and ensure liquid wholesale funding markets. Although the model is based on historical data, it offers insights pertinent to ongoing discussions about

[1]Jerome Powell is the chair of the Federal Reserve. See his full speech here.

[2]"IOR" stands for "interest on reserves," previously also known as "interest on excess reserves" or "IOER."

the Federal Reserve's balance sheet policy and quantitative tightening (QT). Admittedly, due to the limitations of my dataset, the model does not capture every nuanced aspect of monetary policy implementation, but a quantitative framework of this nature, capable of predicting funding rate distortions *out-of-sample*, can be a useful policy tool for central bankers, as it provides valuable guidance for precautionary interventions.

The aggregate quantity of reserves supplied by the Federal Reserve (Fed) is an important policy concern: A sufficient level of reserve balances supports an efficient interbank payment system as well as effective monetary policy implementation (pass-through of the Fed-administered rate into risk-free overnight market rates). An efficient interbank payment system relies on banks making large volumes of timely payments. Because those payments are predominately settled with reserves, a sufficient quantity of reserves "lubricates" the payment system (Atalay, Martin and McAndrews, 2010). Sufficient reserve balances also support the current monetary policy implementation framework, known as "the floor system," by allowing banks to close the spreads between market rates and IOR (Ihrig, Senyuz and Weinbach, 2020): When wholesale market risk-free rates, such as Treasury repo rates, are higher than IOR, banks active in wholesale funding markets ("dealer banks") with extra reserves can in principle lend reserves in wholesale markets, making an arbitrage profit. This forces the spreads between wholesale funding rates and IOR close to zero, thereby enabling the Fed to control market interest rates by adjusting IOR. When reserves are insufficient, however, banks may hoard reserves and not enforce this arbitrage, impeding monetary policy implementation.

In 2019, as total reserve balances decreased,[3] several risk-free funding rates, including Treasury repo rates, sporadically increased above the Fed's target range, as shown in Fig. 1. Notably, on September 16-18, 2019, overnight repo rates spiked more than 300 basis points, attracting considerable attention from global policymakers, academics, and market participants.[4] These spikes surprised the Fed, and necessitated emergency interventions.

[3]Total reserve balances reached a peak of about $2.8 trillion in October 2014. In 2017, the Federal Open Market Committee began implementing balance sheet normalization and planned to reduce its assets and liabilities, including reserves, to the greatest extent consistent with "efficient and effective monetary policy." System-wide reserve balances gradually declined to a low of about $1.4 trillion in early September 2019. See Board of Governors of the Federal Reseve System (2019) for an overview of the Fed's balance sheet normalization policies.

[4]For examples of media reporting, see "Fed Preps Second $75 Billion Blast With Repo Market Still On Edge," *Bloomberg,* September 17, 2019; "Why the U.S. Repo Market Blew Up and How to Fix It," *Bloomberg*, January 6, 2020; "Fed Plans Second Intervention to Ease Funding Squeeze," *Financial Times*, September 17, 2019; "New York Fed Examines Banks' Role in Money Market Turmoil," *Financial Times*, September 20, 2019; "Wall Street Is Buzzing About Repo Rates. Here's Why," *New York Times,* September 18, 2019; "Fed Intervenes to Curb Soaring Short-Term Borrowing Costs,"*Wall Street Journal*, September 17, 2019.

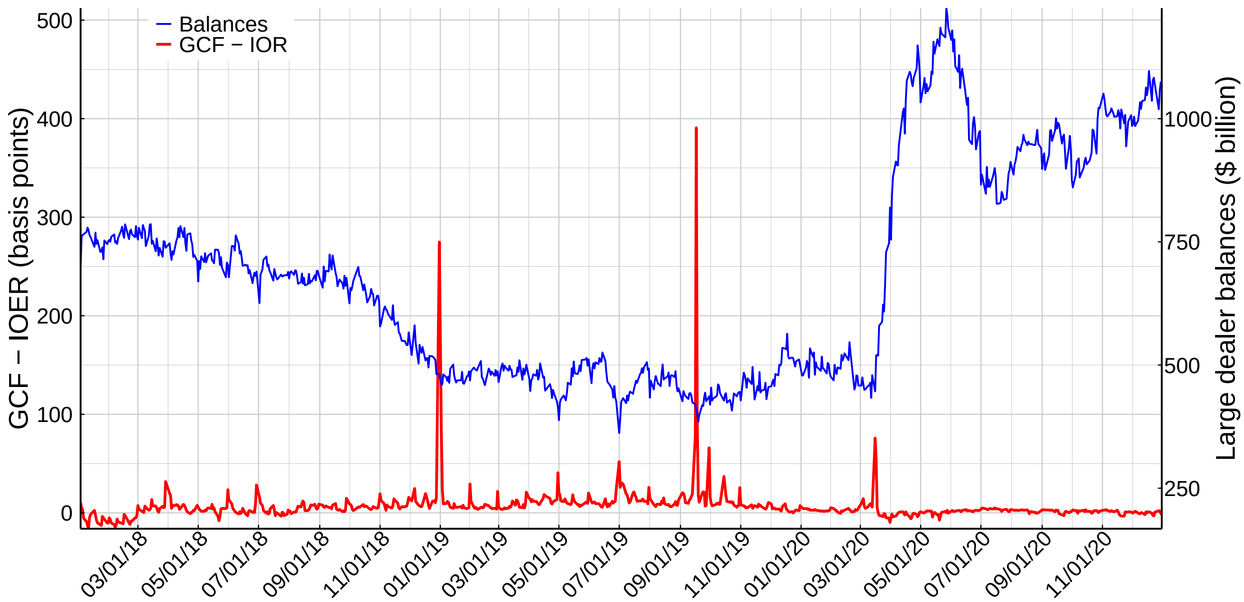

Figure 1: Reserve balances and the spread of the general collateral financing (GCF) repo rate over IOR. GCF is an index of overnight Treasury general collateral interdealer repo rates. IOR, the interest rate paid on reserves, is the Fed's target policy rate. The reserve balances of the large repo-active banks are shown in blue (right axis). The spread of GCF from IOR is shown in red (left axis). Sources: Fedwire Funds Service, FRBNY, Copeland, Duffie and Yang (2020).

The Fed reacted quickly by announcing a series of repo operations and changed the course of its balance sheet normalization process. On September 16-18, 2019, other overnight risk-free funding markets also suffered from liquidity shortages (e.g., the overnight synthetic dollar interest rateas documented by Correa, Du and Liao 2020. See also Fig. 18). These disruptions rippled through various related securities markets (see Figs. 19 to 24). Low reserve balances also reduced the efficiency of the intraday interbank payment system: Copeland, Duffie and Yang (2020) document that the payment timing stress[5] on September 17, 2019 reached its highest level since 2015. The observed link between payment timing stress and funding market spikes is consistent with the mechanism of my paper, which I explain as follows.

My model links two critical empirical facts. First, large dealer banks are the marginal lenders in short-term funding markets, such as repo markets and FX swap markets (Avalos, Ehlers and Eren, 2019; Correa, Du and Liao, 2020). Thus, the equilibrium funding rates are closely related to large dealer banks' marginal value of reserves.[6] Second, in the

[5]Payment timing stress is measured by the payment time net of its sample mean when 50% of the day's total incoming value has been received by the 10 largest dealer banks over Fedwire. The later dealer banks receive their incoming payments, the higher the payment timing stress.

[6]Most transactions in wholesale funding markets are settled using reserve balances at the central bank, so lending in the funding markets causes an outflow of banks' excess reserves (Bech, Martin and McAndrews,

modern interbank payment system, large banks rely heavily on incoming payments from other banks before being able to make the bulk of their own outgoing payments. This reliance gives rise to strategic complementarity in banks' payment timing decisions (Bech and Garratt, 2003; McAndrews and Rajan, 2000; Afonso, Duffie, Rigon and Shin, 2022): For any large bank—say, JPMorgan—incoming payments from other banks provide the balances needed to cover its own outgoing payments, so when JPMorgan believes other banks will send it payments early in the day, it will likewise tend to pay others early. On the other hand, when JPMorgan believes other banks are paying it late—for example, because of low reserve balances—JPMorgan will also pay late.

At the center of my mechanism is the aforementioned strategic complementarity in interbank intraday payment timing. A sufficiently low supply of reserves causes banks to suddenly hoard reserves, reinforced by a feedback effect stemming from the strategic complementarity of intraday payment timing, and leads to intraday payment timing stress. This results in high marginal values of reserves, and disincentivizes banks from efficiently allocating liquidity into wholesale funding markets. As such, when reserves are close to being insufficient, even small reductions in reserve balances can have strong nonlinear or discontinuous impacts on short-term wholesale funding rates, mirroring observed market events such as the wholesale funding rate spikes of September 2019.

To gauge the quantitative implications of this strategic complementarity, I calibrate the parameters of my model using a method-of-moments procedure with data from 2019. My sample comprises days leading up to August 31, 2019. My fitted model is able to predict the funding rate spikes of September 16-18, 2019, as an out-of-sample event (see Fig. 6). This confirms the validity of my model's mechanism and underscores the importance of incorporating strategic complementarity in payment timing when shaping central-bank policy. I then utilize this model to compute a counterfactual estimate of the minimum level of reserves required by large U.S. dealer banks that would have mitigated reserve hoarding and kept the expected spread between GCF repo rates and IOR below 13 basis points (see Fig. 8). To the best of my knowledge, this represents the first estimation of its kind regarding adequate reserve supply in the academic literature under the current policy framework. This quantitative analysis provides insights into the ongoing discourse regarding the Federal Reserve's quantitative tightening (QT) strategy.

2012; Correa et al., 2020).

2 Background

The notion of reserve sufficiency has changed substantially since the GFC. Prior to the 2007-2009 crisis, aggregate reserves typically remained below $50 billion, which was sufficient for both the efficient functioning of wholesale funding markets and for banks to manage their intraday liquidity needs. During this period, the Fed controlled interest rates by actively managing the supply of reserves within the banking system. In contrast, today's post-GFC liquidity requirements mandate that large banks maintain substantial reserve balances at the Fed throughout each day (see Appendix B.3 for more details). Additionally, the level of aggregate reserves has significantly increased due to the Fed's crisis facilities and post-crisis quantitative easing programs. In this new regulatory and macroeconomic environment, the Fed conducts monetary policy by adjusting the overnight interest rate on reserves (IOR) held at the Fed. This approach necessitates that the Fed supplies adequate reserves to dealer banks, yet the level of reserves required to ensure efficient funding markets and a robust interbank payment system remains an unresolved issue.

The funding rate spikes of September 16-18, 2019 were surprising to policymakers and market participants. Lorie Logan, manager of the System Open Market Account (SOMA) for the Federal Open Market Committee (FOMC) at the time, stated on September 20, 2019 , that "*the expectation had been that as repo rates rose, banks would withdraw excess cash held at the Fed and lend it into the repo market... Instead the New York Fed had to step in to provide that cash as banks remained on the sidelines.*" From 2011 to 2019, total excess reserves consistently exceeded $1.4 trillion, significantly surpassing pre-GFC levels. In addition, according to Senior Financial Officer Surveys conducted by the Fed in September 2018 and February 2019 regarding the "lowest comfortable level of reserves," there should still have been ample reserve balances in early 2019 for any single large dealer bank (Andros, Beall, Martinez, Rodrigues, Styczynski and Thorp, 2019).

A body of academic research and some finance industry commentaries underscore the influence of supervisory and regulatory requirements—particularly Basel III liquidity regulations—on reserve sufficiency. These regulations encourage large U.S. banks to maintain significant reserve balances at the Fed throughout each day.[7] These same regulations, however, may have inadvertently weakened monetary policy implementation by inhibiting banks from utilizing these reserves in short-term funding markets (see Anbil, Anderson and Senyuz 2020b, d'Avernas and Vandeweyer 2020, and Nelson and Covas 2019

[7]See more details in Appendix B.3.

among others). However, considering that banks' actual reserves exceeded the levels required by regulations—as indicated by banks regulatory reports and the aforementioned Senior Financial Officer Surveys—one might question the extent to which these regulations actually restricted banks from lending in the repo market in September 2019. Indeed, a NY Fed staff report, Afonso, Cipriani, Copeland, Kovner, La Spada and Martin (2020b), points out that "it seems unlikely that regulation itself may have been a key contributing factor to the money market stress of mid-September. Banks typically hold considerable buffers above their regulatory minima, which means that the regulatory constraints were, in all likelihood, not binding."

My paper addresses the aforementioned puzzle by demonstrating that reserves can suddenly become insufficient at levels well above those required by regulations. Although banks' liquidity positions and stress testing information are routinely available to policymakers, my model reveals that this data alone is inadequate for accurately gauging the minimum aggregate reserve supply necessary for a well-functioning money market and effective monetary policy implementation. Therefore, a comprehensive understanding of how banks utilize reserves for making payments is essential to inform the Federal Reserve's policy regarding reserve supply.

3 Related literature

This paper is closely related to the empirical literature studying financial intermediaries and the mechanism of wholesale funding markets such as the repo market. Large financial intermediaries, especially large U.S. banks, play increasingly important roles in wholesale funding markets. Post-crisis regulations thus have profound implications for wholesale funding markets by influencing intermediaries' balance sheets and liquidity management decisions (Duffie, 2018; Adrian and Shin, 2011; Ranaldo, Schaffner and Vasios, 2020; Egelhof, Martin and Zinsmeister, 2017). Correa, Du and Liao (2020) examine the daily balance sheet information of the large U.S. dealer banks and find they substantially increased the liquidity provision in the FX swap markets and repo markets from 2016 to 2020.[8] In particular, Correa, Du and Liao (2020) point out that post-GFC, key Basel III regulatory ratios such as SLR and GSIB capital surcharge scoring have significantly increased banks' balance sheet costs, so banks heavily rely on draining down their own reserves

[8] Fig. 9 shows the net liquidity provision (lending minus borrowing) of the large U.S. dealer banks to the repo markets, in comparison with the liquidity provision by the money market funds (MMF).

for liquidity provision. [9] Acharya, Chauhan, Rajan and Steffen (2022); Lopez-Salido and Vissing-Jorgensen (2023) study the implications of Quantitative Tightening (QT) on banks' liquidity needs and liquidity provisions respectively.

Concurrent empirical research by Adam Copland, Darrell Duffie, and Yilin (David) Yang investigates the relationship between the total supply of reserves and repo rates. Their key findings, illustrated in Fig. 14 and Table 3, indicate that lower aggregate reserves among large dealer banks correlate with elevated repo rates. Consistent with my theory, Fig. 14 demonstrates that Treasury repo rates as predicted by quantile regressions, becomes increasingly sensitive to fluctuations in dealer balances. Notably, the quantile regression predicts that a one standard deviation decrease in dealer banks' reserves (250 billion dollars) is associated with a rise of up to only 24 basis points in the 99th percentile of the repo rate, which is considerably less than the spikes observed in September 2019. This discrepancy underscores the limitations of reduced-form empirical models, which are incapable of accounting for the feedback effects intrinsic to strategic complementarity, in capturing significant market responses to minor reductions in reserves, highlighting the value of theoretical models in understanding such dynamics. Copeland, Duffie and Yang (2020) also find a strong relationship between interbank intraday payment timing delays and repo rate spikes, as shown in Fig. 15. Additionally, in line with Correa, Du and Liao (2020), Copeland, Duffie and Yang (2020) find evidence supporting the role of demand factors such as Treasury issuance on repo rates. The repo spike of mid-September 2019 is a good example of the interplay between supply and demand factors, detailed comprehensively by Afonso, Cipriani, Copeland, Kovner, La Spada and Martin (2020b), Anbil, Anderson and Senyuz (2020a), Anbil, Anderson and Senyuz (2020b), Ihrig, Senyuz and Weinbach (2020), Avalos, Ehlers and Eren (2019), and Martin, James, Palida and Skeie (2020), among others.

While my work aligns with the above-mentioned empirical research, it stands out as one of the first two papers to illustrate that minor reductions in reserve balances can cause significant, discontinuous increases in wholesale funding rates. In a contemporaneous work, d'Avernas and Vandeweyer (2020) focus on the repo rate spike event of September 2019, constructing a model that examines the impact of binding regulations such as internal Liquidity Stress Tests (LST) on repo rate spikes. They suggest a theoretical possibility that anticipation of future funding market disruptions might have contributed to the un-

[9] Avalos, Ehlers and Eren (2019) show that large U.S. banks have become important net lenders in the repo market since 2011.

expected rise in Treasury spreads in March 2020. In contrast, my paper highlights the central role of strategic complementarity in banks' intraday payment timing on wholesale funding rate spikes and delves into reserve sufficiency from the perspectives of an efficient interbank payment system and effective monetary policy implementation.

My paper is related to the literature studying global games and their applications in finance. In my paper, the central mechanism is the strategic complementarity in banks' payment-timing decisions. The techniques of global games that incorporate strategic complementarity in various forms are extensively applied to examine phenomena such as currency attacks, bank runs, debt crises, and the determination of safe assets (Carlsson and Van Damme, 1993; Morris and Shin, 1998, 2001; Goldstein and Pauzner, 2005; He and Xiong, 2012; Heider, Hoerova and Holthausen, 2015; He, Krishnamurthy and Milbradt, 2019; Liu, 2016). A distinctive feature of my model, compared to standard global games models, is that agents' beliefs about the payoff-relevant states are endogenously determined in equilibrium. In this context, my model contrasts sharply with the findings of Angeletos, Hellwig and Pavan (2006) and Angeletos and Werning (2006), who suggest that endogenously generated information can lead to multiple equilibria in global games. Conversely, under mild technical conditions, my model guarantees a unique equilibrium. This uniqueness arises partly from the interaction between strategic substitutability and complementarity within my model, a feature elaborated in Section 4.2. Such interactions are novel and relatively underexplored in the literature, and introduce several new challenges to my model. Due to these challenges, standard techniques from global games are not directly applicable to my setting; instead, I have devised a novel proof method to characterize the equilibrium.

A strand of literature explores banks' liquidity hoarding in the overnight funding markets, attributing this behavior to adverse selection and counterparty risk (Heider, Hoerova and Holthausen, 2015; Gorton and Metrick, 2012; Acharya and Merrouche, 2013). In contrast, my model attributes precautionary hoarding of reserves in interbank intraday payment systems to strategic complementarity.

Copeland, Duffie and Yang (2020) and this paper are the first to point out the close connection between the intraday interbank payment mechanism and monetary policy implementation in the post-GFC regulatory environment. Due to financial-stability concerns, central banks and scholars worldwide have extensively studied reserve abundance and intraday payment mechanisms among banks (see Afonso, Duffie, Rigon and Shin (2022); Armantier, McAndrews and Arnold (2008); McAndrews and Rajan (2000); Bech

(2008); Bech and Garratt (2003); Schoenmaker (1995); Zhou (2000); McAndrews and Potter (2002); McAndrews and Rajan (2000) among others). The recent study by Goldstein, Yang and Zeng (2023) examines the interbank payment system using tools from the repeated games literature, and also explore equilibrium uniqueness in their setting. There is also a substantial body of literature on monetary policy implementation with an ample supply of reserves (Logan, 2019; Afonso, Kim, Martin, Nosal, Potter and Schulhofer-Wohl, 2020a; Piazzesi, Rogers and Schneider, 2019; Lenel, Piazzesi and Schneider, 2019); however, intraday payment mechanisms were not previously considered relevant to monetary policy implementation. This paper highlights that, following the introduction of post-crisis liquidity regulations, banks have become reluctant to use the Fed's intraday overdraft facility, making their own reserves the most crucial liquidity source. Consequently, the effects of strategic complementarity have intensified post-GFC, under current macroeconomic conditions, forging a strong link between the intraday payment system and monetary policy implementation.

The rest of the paper is organized as follows: Section 4 introduces the baseline model. Section 5 characterizes equilibrium under different macroeconomic conditions. Section 6 explores how minor reductions in reserves can lead to sudden liquidity crunches, offering a theoretical explanation for the repo spike in September 2019. Section 7 outlines the data utilized in this study. Section 8 extends and calibrates the model using method of moments. Section 9 presents the main findings: a counterfactual analysis of the sufficient level of reserves necessary to support effective monetary policy implementation and an efficient interbank payment system. Section 10 provides the concluding remarks.

4 Baseline model

There are two types of agents: n dealer banks active in liquidity provision in the overnight USD wholesale funding markets[10] and n overnight wholesale funding borrowers. The timing of the model is shown in Fig. 2. Each business day is divided into four time periods: 0, 1, 2, and 3. Initially, at time 0, bank i observes only R_i, its beginning reserve balance, and borrower i observes only D_i, its financing target to borrow. The day opens

[10] One prominent feature of the USD wholesale funding markets is the outsized importance of a few large U.S. banks. Copeland, Duffie and Yang (2020) focus on 10 large repo-active dealer banks and show the total reserve balances of large financial institutions outside these 10 were much less influential with respect to repo rates. Similarly, Correa, Du and Liao (2020) study six global systemically important banks (GSIBs)—Bank of America, Citi, Goldman Sachs, JP Morgan, Morgan Stanley, and Wells Fargo—and show they are major liquidity providers in both the repo markets and the FX swap markets.

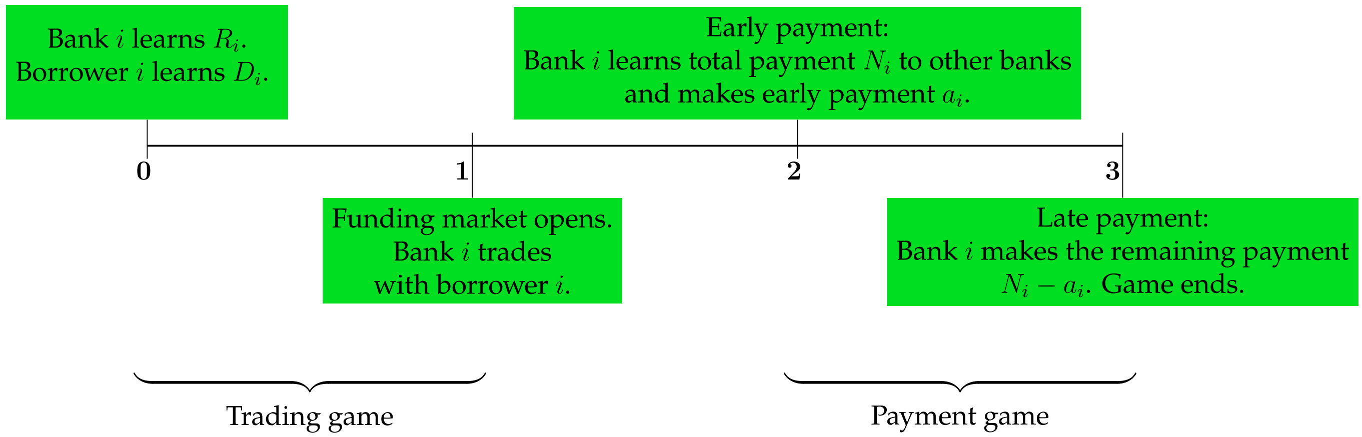

Figure 2: The timeline of the model.

with trading in the wholesale funding markets at time 1. In this market, bank i and borrower i play the trading game. For simplicity, I assume that whenever bank i lends an amount, say S_i, to borrower i in the overnight wholesale funding markets, bank i will transfer S_i quantity of reserves to the clearing bank of borrower i.

After making an overnight loan S_i in the funding markets, at time $t = 2$, bank i observes the total customer payment obligation N_i that bank i must send to other banks by the end of the business day ($t = 3$). Subsequently, the banks play a payment game. In this subgame, bank i can choose to pay any smaller amount a_i early to the other banks at t=2, and delivers the reminder $N_i - a_i$ later at $t = 3$.

The timeline is broadly consistent with the operational details of the USD wholesale funding markets and the interbank payment system, specifically the Fedwire. For instance, in the repo market—the principal wholesale funding markets—approximately 90% of the trading volume is completed before 9 a.m. (EST), as documented by Copeland, Duffie and Yang (2020). In contrast, about 50% of interbank intraday payments are not completed before 1 p.m. (EST), as shown by Copeland, Molloy and Tarascina (2019). (To fix ideas, consider $t = 1$ representing the period from 6 a.m. to 9 a.m., $t = 2$ from 9 a.m. to 2 p.m., and $t = 3$ from 2 p.m. to 7 p.m.)

To clarify the main ideas, I assume there are only two banks (bank 1 and bank 2) and 2 borrowers (borrower 1 and borrower 2) in the baseline model.[11] The random variables N_1 and N_2 are assumed to be identically and independently distributed according to a cumulative distribution function (cdf) $F_N(\cdot)$, which has a finite mean and supports $[N_{min}, \infty)$, where $N_{min} > 0$. The function $F_N(\cdot)$ is atomless and strictly increasing over

[11]See Appendix D.2 for a general model with $n > 2$ banks and borrowers.

(N_{min}, ∞). Additionally, R_1 and R_2, as well as D_1 and D_2, are assumed to be identically and independently distributed with their respective cdfs $F_R(\cdot)$ and $F_D(\cdot)$. For simplification, I assume these variables are independent, though my results extend to arbitrary joint distributions over (R_1, R_2, D_1, D_2). The distributions of all random variables—R_1, R_2, D_1, D_2, N_1, and N_2—are common knowledge, but their realizations are private information. The model is described from the perspective of bank i, with the other bank denoted as bank j. The setups and results are symmetric for both bank i and bank j.

4.1 Payment subgame from $t = 2$

In the modern interbank payment system, settlement of time-critical interbank payments relies on reserves (McAndrews and Kroeger, 2016; Soramäki, Bech, Arnold, Glass and Beyeler, 2007). The volume of payment requests far exceeds the amount of total reserves in the banking system, as shown in Fig. 11. Large banks have to rely heavily on incoming payments from other banks before being able to make the bulk of their own outgoing payments, and therefore face a serious liquidity management problem when customers' payment requests outbalance incoming payment flows, as detailed in Appendix B.2. From the lens of my model, this means that N_i is likely to be much larger than R_i.

As documented by Bech, Martin and McAndrews (2012), once bank i transfers the overnight loan S_i to the clearing bank of borrower i, that those reserves becomes unavailable for settling customer payment requests. After observing N_i, bank i can pay any positive amount $a_i \leq N_i$ at time 2, deferring the remaining payments until time 3. The amount a_i must be measurable with respect to the bank i's information set after trading in the funding market. This information is fully represented by the reserve balances L_i of bank i after trading, net of the minimum required reserve balances of bank i before time 2. That is, $L_i = R_i - S_i - Q$, where $Q > 0$ represents an exogenous constant, denoting the minimum level of reserves mandated by liquidity regulations.

It is worth mentioning that in the pre-crisis era, absent those liquidity regulations, banks had often kept minimum reserve balances and relied heavily on borrowing from the Fed's intraday overdraft facility to meet ongoing payment demands (Fig. 13). Borrowing from the Fed, however, signaled to bank supervisors that banks have negative intraday reserve balances.[12] By contrast, the post-GFC new liquidity requirements incentivized large

[12]Large U.S. banks can in principle borrow additional reserves from other financial institutions, for example, from money market funds in the Tri-party repo market, to make outgoing payments. However, borrowing reserves increases the supplementary leverage ratio (SLR)—a more binding constraint—of these

U.S. dealer banks to maintain substantial balances at the Fed and discouraged them from incurring daylight overdrafts on their reserve accounts,[13] highlighting a stark difference from the pre-GFC era where such mechanisms played a less significant role in monetary policy implementation.

After the payments at time 1, $L_i - a_i + a_j$ is the residual reserve balance of bank i. Bank i bears a cost if this residual balance is below 0, modeled as a per-unit "regulatory cost" of $\psi > 0$. In reality, banks are especially worried about the cost of a failure to satisfy strict new supervision of liquidity sufficiency (Correa, Du and Liao, 2020; d'Avernas and Vandeweyer, 2020), and the stigma in the eyes of their supervisors associated with borrowing from the Fed's intraday overdraft facility. This implies that ψ is very large. Bank i also suffers a linear cost $c \cdot (N_i - a_i)^+$ caused by paying bank j late, for some late payment cost coefficient $c > 0$. (I call c the "marginal cost of delay.") Costs associated with late payments are discussed extensively in the literature studying banks' intraday liquidity management, including work by Ashcraft, McAndrews and Skeie (2011), Afonso and Shin (2011), Bech and Garratt (2003), and Bech (2008). The final cost to bank i associated with payment timing is thus

$$\underbrace{\psi \cdot (L_i - a_i + a_j)^-}_{\text{liquidity cost}} + \underbrace{c \cdot (N_i - a_i)^+}_{\text{late payment cost}}.$$

Here, I adopt the convention that

$$(L_i - a_i + a_j)^- = \begin{cases} 0 & \text{if } L_i - a_i + a_j \geq 0, \\ |L_i - a_i + a_j| & \text{if } L_i - a_i + a_j < 0. \end{cases}$$

The cost function to bank i captures the force of strategic complementarity in banks' payment timing decisions (Afonso and Shin, 2011). That is, the higher the early payment strategy a_j of bank j conjectured by bank i, the higher the best-response early payment a_i of bank i. Given the payment strategy a_j of bank j, bank i chooses a_i to optimize the conditional expected payoff,

$$U(L_i, N_i) = \mathbb{E}[-\psi\,(L_i - a_i + a_j)^- - c\,(N_i - a_i)^+ \mid N_i, L_i]. \tag{1}$$

Lemma 1. *Suppose L_i and L_j are arbitrarily distributed such that $\mathbb{P}(L_i > 0) > 0$ and $\mathbb{P}(L_j > 0) > 0$. Then there is a Perfect Bayes payment game equilibrium of the form $a_i^* = \min((L_i + \alpha_i)^+, N_i)$,*

banks, and the Tri-party market typically settles late in the day, from 3 p.m. to 6 p.m. EST. Consequently, borrowing reserves is not particularly useful for the purpose of making interbank payments. For simplicity, I assume banks cannot borrow reserves at time 2.

[13] This is corroborated by conversations with senior managers at several Global Systemically Important Banks (GSIBs), as documented by Copeland, Duffie and Yang (2020).

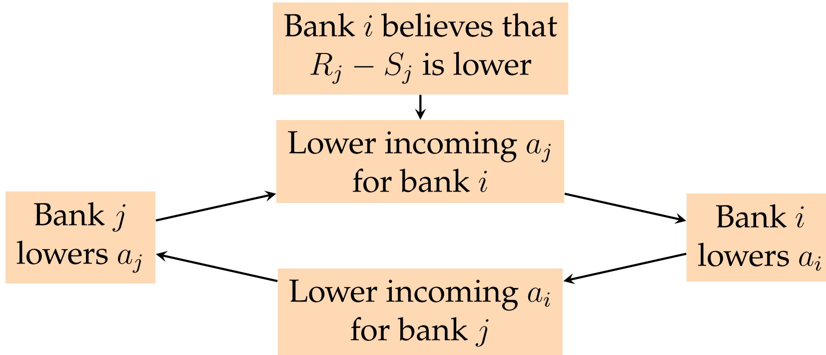

Figure 3: Intuition for Lemma 1. A "vicious circle" causes banks to hoard reserves.

$a_j^* = \min((L_i + \alpha_j)^+, N_j)$ *for some constants* α_i, α_j *such that*

$$\alpha_i = \inf\left\{ \vartheta \geq 0: \quad \mathbb{P}(\min((L_j + \alpha_j)^+, N_j) \leq \vartheta) \geq \frac{c}{\psi} \right\}$$
$$\alpha_j = \inf\left\{ \vartheta \geq 0: \quad \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \vartheta) \geq \frac{c}{\psi} \right\}.$$

In particular, when L_i *and* L_j *have the same distribution*[14] *, then* $\alpha_i = \alpha_j = \alpha$*. If* $\mathbb{P}(L_j \leq 0) > c/\psi$*, then* $\alpha = 0$*. If* $\mathbb{P}(L_j \leq 0) \leq c/\psi$*, then* $\alpha \geq N_{min}$ *and*

$$\alpha = \inf\left\{ \vartheta \geq 0: \ \mathbb{P}(N_j \leq \vartheta) \geq \frac{\frac{c}{\psi} - \mathbb{P}(L_j \leq 0)}{1 - \mathbb{P}(L_j \leq 0)} \right\}. \tag{2}$$

The equilibrium is unique, except for the knife-edge case where $\mathbb{P}(L_j \leq 0) = \mathbb{P}(L_i \leq 0) = c/\psi$*.*

Proofs of all results, including Lemma 1, are in Appendix F. Intuitively, when $\mathbb{P}(L_j \leq 0)$ is high, bank i anticipates that bank j will not have sufficient liquidity to make early payments. Consequently, bank i becomes conservative about paying bank j early. Expecting a lower a_i, bank j reasons along the same lines, leading to a lower a_j. Due to strategic complementarity, this caution reinforces itself: when bank i expects a_j to be lower, it is incentivized to further lower a_i and so on, initiating a "vicious circle." Ultimately, when $\mathbb{P}(L_j \leq 0)$ exceeds a certain threshold, both banks begin to hoard reserves, resulting in $\alpha = 0$. This intuition is illustrated in Fig. 3.

The force that makes the equilibrium unique is similar to the force in the standard global games literature (Carlsson and Van Damme, 1993; Morris and Shin, 1998, 2001; Goldstein and Pauzner, 2005). Here, the strategic complementarity in banks' payment

[14]It turns out that on the equilibrium path, L_i and L_j will always have the same distribution, which will be proved by Theorems 1 and 2.

strategies and the uncertainty of the other bank's liquidity situation L_j and payment obligation N_j lead bank i to make the risk dominant payment action: $a_i = \min((L_i+\alpha_i)^+, N_i)$.[15] However, it is worthwhile highlighting the difference between my model and the standard global games models. In those models, agents receive *exogenous* private signals of the payoff-relevant states. Conversely, in my model, the payoff-relevant states L_i, L_j are *endogenously* determined by the trading outcome in the funding market: $L_i = R_i - Q - S_i$. Hence, even though there is a unique equilibrium for the payment subgame, this does not automatically imply that the entire game has a unique equilibrium (Angeletos, Hellwig and Pavan, 2006; Angeletos and Werning, 2006).

To derive the banks' optimal lending decisions in the funding market, I first need to characterize the banks' marginal value of reserves for the payment subgame. The continuation value of bank i for reserve balances at the beginning of the payment game, before observing its payment obligation N_i, is

$$V(L_i) = \mathbb{E}[U(L_i, N_i) \mid L_i]. \tag{3}$$

When well defined, the left-hand derivative of the value function, as specified in (3), at a given reserve balance y is denoted by $V'_-(y) = \lim_{x \uparrow y} \frac{V(x)-V(y)}{x-y}$.

Lemma 2. *Suppose that in the payment game, bank j makes payment $a_j = \min((L_j + \alpha_j)^+, N_i)$ and bank i makes payment $a_i = \min((L_i+\alpha_i)^+, N_i)$, for some constants α_j, α_i. When $L_i+\alpha_i > 0$, then for bank i,*

$$V'_-(L_i) = \int_{n \in [L_i^+, (L_i+\alpha_i))} \psi \mathbb{P}(a_j \le n - L_i)\, dF_N(n) + \int_{n \in [(L_i+\alpha_i), \infty)} c dF_N(n).$$

When $L_i + \alpha_i \le 0$, for bank i,

$$V'_-(L_i) = \psi \mathbb{P}(a_j \le -L_i).$$

Also, $V'_-(\cdot)$ weakly decreases.

Lemma 2 guarantees the existence of the left-hand derivative of the value function, although $V(\cdot)$ may not be differentiable. Note that $V'_-(\cdot)$ for bank i depends on the payment strategy a_i and a_j. To make this dependence relationship explicit, I define the marginal value of liquidity function $\Gamma_i = V'_-(L_i)$ for bank i as follows:

Definition 1. Fix any payment strategy $a_i = \min((L_i + \alpha_i)^+, N_i)$ and the probability distribution of L_i for all $i \in \{1, 2\}$. Let $j \in \{1, 2\} \setminus \{i\}$. Define the *marginal value of*

[15] See Morris and Shin (2001) for a discussion of the risk dominant strategy.

liquidity function for bank i to be the function $\Gamma_i : \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\Gamma_i(y, \alpha_i, \alpha_j) = \begin{cases} \int_{n \in [y^+, (y+\alpha_i))} \psi \mathbb{P}(\min((L_j + \alpha_j)^+, N_j) \le n - y)\, dF_N(n) + \int_{n \in [(y+\alpha_i), \infty)} c\, dF_N(n), & \text{if } y > -\alpha_i \\ \psi \mathbb{P}(\min((L_j + \alpha_j)^+, N_j) \le -y), & \text{if } y \le -\alpha_i \end{cases}$$

The function $\Gamma_i(\cdot)$ naturally arises when large dealer banks optimize the quantity of reserves to lend in the funding market. The specifics of this process will be elaborated in the following section.

4.2 Trading game at $t = 1$

In this paper, I focus on the overnight wholesale funding markets, although my model can be extended to study other short-term funding markets as well. I model the overnight wholesale funding markets as over-the-counter markets where borrower i is matched with bank i, and they trade bilaterally. At the beginning of each business day, borrower i targets a total borrowing demand of D_i. The net cost of borrower i associated with financing some amount q at some funding rate r (endogenously determined in equilibrium) is[16]

$$qr + \underbrace{\frac{\xi}{2}((D_i - q)^+)^2}_{\text{cost of reduced financing}},$$

where the marginal value of financing, ξ, is a positive coefficient determining the sensitivity of the cost to borrower i of unmet financing needs. I assume $D_i > D_{min}$ almost surely for a constant D_{min}, and $\mathbb{P}(R_i - Q - D_{min} > 0) > 0$.[17]

The exact form of the equilibrium funding rate r largely depends on the market microstructure of the funding markets, such as the bargaining power between borrowers and lenders. Nevertheless, the qualitative predictions of my main results remain robust across various market microstructure configurations, provided that the funding rate is an increasing function of $\Gamma_i(y, \alpha_i, \alpha_j)$. This is not surprising, because the driving force is the aforementioned strategic complementarity in the payment subgame. I discuss results for one concrete example of competitive pricing in the baseline model, where both bank i

[16]For simplicity, I normalize the Fed's policy rate benchmark, IOR, to be zero. When IOR is not zero, r should be understood as the spread between the gross funding rates r^{gross} and IOR, and the cost for borrower i should be adjusted to $qr^{gross} + (D_i - q)^+ \cdot \text{IOR} + \frac{\xi}{2}((D_i - q)^+)^2 = qr + \frac{\xi}{2}((D_i - q)^+)^2 + D_i \cdot \text{IOR}$ (in equilibrium $q < D_i$). The rest of my analysis does not change.

[17]Empirically, banks never have negative reserve balances before they start making interbank intraday payments.

and borrower i behave as price takers. In Appendix D.1, I include results under another funding market structure—the case of monopolistic pricing, where bank i acts as a local monopolist by offering a supply schedule $g : \mathbb{R} \to \mathbb{R}$ to screen the borrower's demand and maximize profit.

For the baseline model, I assume that bank i and borrower i are fully competitive. Bank i takes the overnight funding rate r_i as given and lends the quantity s, solving

$$\sup_s \; sr_i + V(R_i - s),$$

where sr_i is the interest paid by borrower i to bank i on the next business day. Borrower i minimizes net cost by solving

$$\inf_s \; sr_i + \tfrac{\xi}{2}((D_i - s)^+)^2.$$

The first-order conditions for bank i and borrower i imply the equilibrium funding rate r_i^* and quantity S_i^* satisfy

$$\begin{aligned} S_i^* &= \inf \Big\{ s : \Gamma_i(R_i - Q - s, \alpha_i, \alpha_j) \geq \xi(D_i - s) \Big\} \\ r_i^* &= \xi(D_i - S_i^*), \end{aligned} \tag{4}$$

where the marginal value of liquidity Γ_i is defined in Definition 1. Clearly, ξ governs the demand elasticity.

Throughout the paper, the equilibrium concept is a perfect Bayesian equilibrium. By definition, the distribution of $L_i = R_i - Q - S_i$ and $L_j = R_j - Q - S_j$ is endogenously determined by the equilibrium outcome of the trading game. From Eq. (4), the equilibrium outcome of the trading game depends on Γ_i and Γ_j, which in turn depend on the distribution of $L_i = R_i - Q - S_i$ and $L_j = R_j - Q - S_j$ (Lemma 1). Thus, characterizing the full equilibrium will necessarily involve solving a complicated fixed-point problem. In particular, there is no a priori reason why the model has a unique equilibrium.

It is worthwhile to highlight the presence of strategic *substitutability* in banks' lending decisions, which interacts with the aforementioned strategic complementarity. This dual interaction is rarely explored and introduces several new complications. For instance, when bank i perceives that bank j will choose an equilibrium strategy to offer a smaller loan S_j in the funding market, it infers that bank j will retain more reserves, $L_j = R_j - S_j - Q$, and therefore could afford a larger payment a_j in the payment subgame. A larger a_j reduces the bank i's marginal value of reserves for the subgame. Consequently, bank i is incentivized to increase its own lending amount S_i in the funding market. This mechanism implies that, even if R_i and R_j share the same probability distribution, the endogenous distributions of L_i and L_j in equilibrium may not be symmetrical. At first glance, it appears that the equilibrium of the entire game might depend on an infinite hierarchy of beliefs between the two banks: bank i's decision to lend S_i is contingent upon its beliefs about bank j's strategy, such as α_j, which in turn depends on bank j's belief about bank i's strategy, and so forth.

5 Liquidity stress index and equilibrium outcome

Any change in macroeconomic conditions, such as a reduction in reserve balances, an increase in borrowing demand in the funding markets, or change in the regulatory requirements, may change the distributions of R_i, D_i and the value of Q. Those changes will eventually change banks' beliefs, their strategies in the payment subgame, and the equilibrium funding rates. Nevertheless, I demonstrate that these complicated macroeconomic conditions can be summarized by one index—the *liquidity stress index*.

Definition 2. Let F_R be the cdf of R_j. The *liquidity stress index* is

$$m \equiv \mathbb{P}\left(R_j - D_j - Q \leq -\frac{c}{\xi}\right) - \frac{c}{\psi} = \mathbb{E}\left[F_R\left(D_j + Q - \frac{c}{\xi}\right)\right] - \frac{c}{\psi},$$

The *liquidity hoarding condition* is when

$$m = \mathbb{E}\left[F_R\left(D_j + Q - \frac{c}{\xi}\right)\right] - \frac{c}{\psi} > 0. \tag{5}$$

The *no hoarding condition* is when

$$m = \mathbb{E}\left[F_R\left(D_j + Q - \frac{c}{\xi}\right)\right] - \frac{c}{\psi} < 0. \tag{6}$$

The liquidity hoarding condition applies whenever the probability distribution F_R of initial reserve balances is sufficiently low in the sense of first order stochastic dominance, for given parameters Q, c, ξ, ψ, D_{min}. Let $F_{RD}(y) = \mathbb{P}(R_j - D_j - Q \leq y)$. The next two key theorems elucidate how the liquidity stress index determines the level of reserve hoarding in equilibrium. Fig. 4 illustrates the intuition for the mechanism of the these results.

Theorem 1. *Under the liquidity hoarding condition, there is a unique equilibrium. In this equilibrium, bank i hoards reserves and pays $a_i^* = \min(L_i^+, N_i)$ in the payment subgame. The marginal value of liquidity functions are the same for both banks $\Gamma_i(y,0,0) = \Gamma_j(y,0,0) = \Gamma(y,0,0)$ and are characterized as follows:*

1. *When* $y > 0$, $\Gamma(y, 0, 0) = c(1 - F_N(y))$;

2. *When* $y \leq 0$, $\Gamma(y, 0, 0) = \psi\left(F_N(-y) + (1 - F_N(-y))F_{RD}(-y - \frac{\Gamma(-y,0,0)}{\xi})\right)$.

Theorem 2. *Under no hoarding condition, there always exists at least one equilibrium. Any equilibrium must be symmetric (in the sense that* $\alpha_i = \alpha_j = \alpha$*) with pure payment strategy* $a_i^* = \min(N_i, (L_i + \alpha)^+)$ *for some* $\alpha > N_{min}$*. The marginal value of liquidity functions for both banks (*$\Gamma_i(y, \alpha, \alpha)$ *and* $\Gamma_j(y, \alpha, \alpha)$*) are identical, hence denoted by* $\Gamma(y, \alpha, \alpha)$*. Furthermore,* α *and* Γ *solve a system of integral equations:*

$$\mathbb{P}\left(R_i - D_i - Q \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right) + \mathbb{P}(N_i \leq \alpha)\left(1 - \mathbb{P}\left(R_i - D_i - Q \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)\right) = \frac{c}{\psi};$$

$$\Gamma(y,\alpha,\alpha) = \begin{cases} \psi \displaystyle\int_{n \in (y^+,(y+\alpha))} F_N(n-y) + (1 - F_N(n-y))F_{RD}(n - y - \alpha - \tfrac{\Gamma(n-y-\alpha,\alpha,\alpha)}{\xi})dF_N(n) & \\ \qquad + \displaystyle\int_{n\in[(y+\alpha),\infty)} c\, dF_N(n), & \forall y > -\alpha; \\ \psi\left(F_N(-y) + (1 - F_N(-y))F_{RD}(-y - \alpha - \tfrac{\Gamma(-y-\alpha,\alpha,\alpha)}{\xi})\right) & \forall y \leq -\alpha. \end{cases}$$

Eq. (4) and Theorems 1 and 2 together imply that the equilibrium funding rate depends on (1) initial reserve balances R_i of bank i, (2) equilibrium trading quantity S_i[18] and (3) liquidity stress index m.

Whereas Theorem 2 does not show equilibrium uniqueness under the general distributional assumptions for the exogenous state variables, it does characterize all *possible* equilibria by a system of non-standard integral equations. In particular, all equilibria share the same prediction: the funding rate spikes when the liquidity stress index turns from negative to positive (see Theorem 4, Theorem 5, and Fig. 5). The next theorem proves the equilibrium is unique under some technical conditions.

Theorem 3. *Assume* $N_i - N_{min}$ $(i = 1, 2)$ *is exponentially distributed with parameter* λ_N *and* F_{RD} *is differentiable with density function* f_{RD}*. Let* $f_{RD}^m = \sup\{f_{RD}(t) : t \leq 0\}$*. If* $\frac{\sqrt{2}e\xi}{\psi} > f_{RD}^m$*, the equilibrium is unique under the no hoarding condition.*[19]

[18] It can be easily shown that conditional on $R_i = \zeta$, there is a monotone relationship between D_i and the equilibrium trading quantity S_i.

[19] The assumptions for Theorem 3 are likely to hold. Realistically, the marginal value of financing ξ cannot be too small relative to the marginal cost of delay c (as confirmed by quantitative estimates from my sample). Recall that under the no hoarding condition, $F_{RD}(\frac{c}{\xi}) < \frac{c}{\psi} < \frac{\sqrt{2}e\xi}{\psi}$, so $\frac{\sqrt{2}e\xi}{\psi} > f_{RD}^m$ is generally satisfied for most common families of probability density functions. Moreover, if bank i has better information about the other bank's reserve condition, that is, uncertainty about $R_j - D_j - Q$ is smaller, f_{RD}^m is smaller. Therefore, the equilibrium is unique, especially when banks have more precise information about each other.

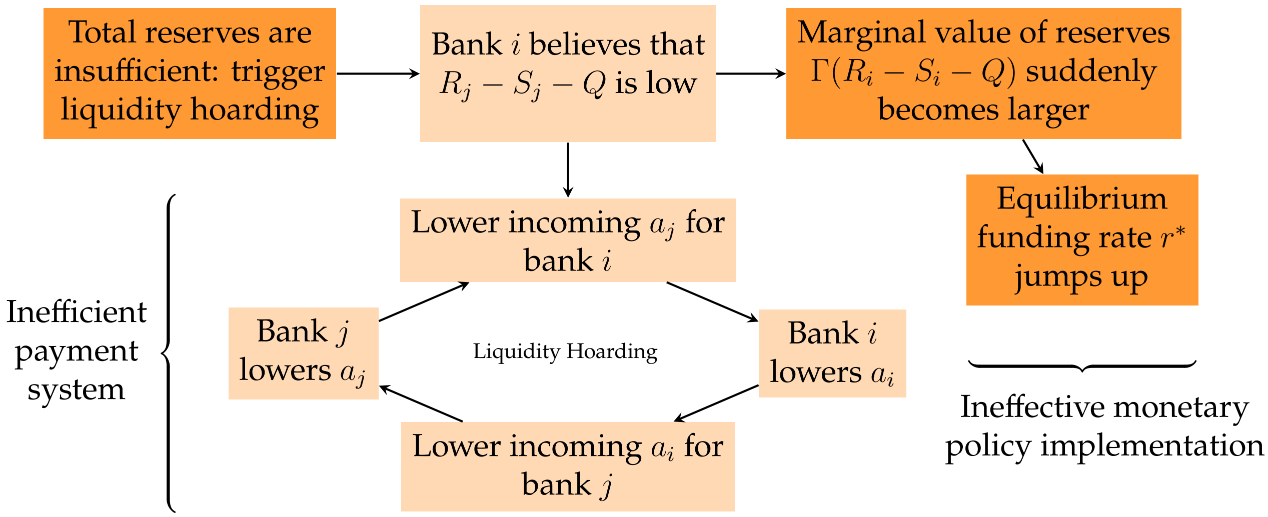

Figure 4: Low supply of reserves causes ineffective monetary policy implementation and an inefficient interbank payment system.

6 How small reductions in reserves trigger liquidity crunches

It is not surprising that significant shifts in macroeconomic conditions can substantially influence banks' equilibrium payment timing decisions and equilibrium wholesale funding rates. Nonetheless, I demonstrate that even small or moderate changes in these conditions—in particular, minor reductions in *reserve supply*—can elicit strong nonlinear or discontinuous impacts on wholesale funding rates, as evidenced by the events of September 2019. To substantiate this claim, it is essential to formalize the concept of "small changes" between two macroeconomic conditions. To begin, in this model, all exogenous constants and functions together represent a macroeconomic condition.

Definition 3. A *set of macroeconomic conditions*, $\mathcal{M}_C$, is comprised of the constants and probability distributions of the exogenous random variables in this economy, denoted by $\mathcal{M}_C = \{F_R(\cdot), F_N(\cdot), Q, \lambda, D_{min}, c, \psi, \xi\}$.

Generically, any minor alterations in macroeconomic conditions are likely to induce slight changes in the liquidity stress index. Given that the liquidity stress index serves as the sufficient statistic for determining equilibrium states, it is intuitive to measure the proximity between two macroeconomic conditions in terms of differences in this index. To formalize this idea, I propose the following definition of closeness between macroeconomic conditions:

Definition 4. We say that any two sets of macroeconomic conditions, denoted by $\mathcal{M}_C^1 = \{F_R^1(\cdot), F_N^1(\cdot), Q^1, \lambda^1, D_{min}^1, c^1, \psi^1, \xi^1\}$ and $\mathcal{M}_C^2 = \{F_R^2(\cdot), F_N^2(\cdot), Q^2, \lambda^2, D_{min}^2, c^2, \psi^2, \xi^2\}$, are *close with respect to the liquidity stress index* if there is some constant $\mathcal{O}$ such that[20]

$$\sup\{\ |c^1 - c^2|,\ |\psi^1 - \psi^2|,\ |\xi^1 - \xi^2|,\ |\lambda^1 - \lambda^2|,\ |D_{min}^1 - D_{min}^2|,\ ||F_{RD}^1 - F_{RD}^2||_\infty\ \} < \mathcal{O}|m^1 - m^2|,$$
$$||F_N^1 - F_N^2||_\infty = 0,$$

where m^1 and m^2 are the liquidity stress indexes under $\mathcal{M}_C^1$ and $\mathcal{M}_C^2$, respectively.

By Definition 4, if $\mathcal{M}_C^1$ and $\mathcal{M}_C^2$ are close with respect to the liquidity stress index, and $|m_1 - m_2|$ is small, then the distances between any elements from $\mathcal{M}_C^1$ and $\mathcal{M}_C^2$ are also close to zero. For ease of reference and with a slight abuse of notation, macroeconomic conditions will be indexed by their liquidity stress index values. For example, $\mathcal{M}_C^1$ will be referred to as $\mathcal{M}_C^{m_1}$.

Aggregate reserves steadily yet slowly declined from March 2017 to September 2019 under the Fed's balance sheet normalization policy. Day-to-day variations in macroeconomic conditions during this period were minimal. Specifically, there were no significant shocks between Friday, September 13, and Monday, September 16, 2019.[21] Nevertheless, even these minor differences were sufficient to trigger the liquidity hoarding condition, significantly impacting the equilibrium wholesale funding rates, as demonstrated by Theorems 4 and 5, providing an explanation for the abrupt spike in repo rates on September 16.

Theorem 4. *Fix some realization ζ of beginning reserve balances R_i and a quantity $\mathcal{S}^*$ traded in the funding market such that $\zeta - \mathcal{S}^* \neq Q$. The equilibrium funding rate r^* jumps up as a function of the liquidity stress m at the threshold $m = 0$ that triggers liquidity hoarding. More specifically, there exists some $\delta(\zeta, \mathcal{S}^*) > 0$ such that*

$$\lim_{\epsilon^m \downarrow 0} r^*(\zeta, \mathcal{S}^*, \epsilon^m) - r^*(\zeta, \mathcal{S}^*, -\epsilon^m) > \delta(\zeta, \mathcal{S}^*),$$

provided that the sets of macroeconomic conditions ${\mathcal{M}_C}^{\epsilon^m}$ and ${\mathcal{M}_C}^{-\epsilon^m}$ are mutually close with respect to the liquidity stress index.[22]

[20] As usual, $||\cdot||_\infty$ is the sup-norm: $\|f\|_\infty = \sup\{|f(x)|\}$

[21] Two factors—corporate tax payments and Treasury issuances—are often cited as explanations for the repo rate spike in September 2019. However, the combined total of these factors between September 16 and 18, 2019, did not rank among the top ten largest for the year and was less than half the size of those on January 31, 2019.

[22] The equilibrium funding rate function in this theorem is $r^* : \mathbb{R}^3 \to \mathbb{R}^+$, which determines the equilibrium supply curve: given some beginning reserve balances $R_i = \zeta$ and the liquidity stress index m that

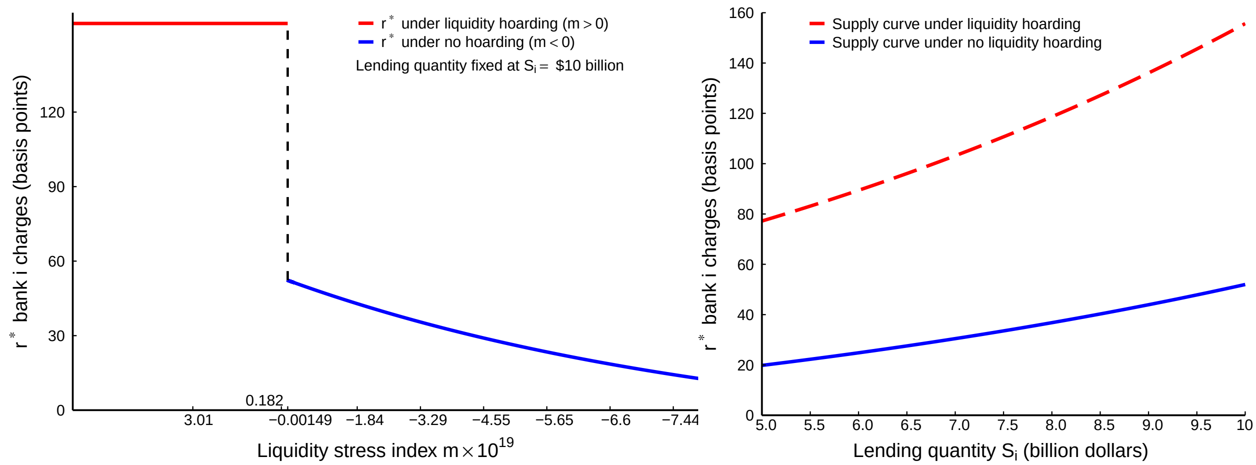

Figure 5: Effects of changes in the liquidity stress index on the funding rate charged by bank i. Both panels fix bank i's beginning reserves at \$40 billion. In the left panel, from right to left, the x-axis represents the increase of liquidity stress index as bank j's reserves decreases from \$40 billion to \$20 billion. The lending quantity S_i remains fixed at \$10 billion in the left panel. This panel illustrates how slight changes in the liquidity stress index can lead bank i to charge significantly higher rates. In the right panel, the x-axis represents the lending quantity and displays bank i's liquidity supply curves: the top red dashed curve is under liquidity hoarding conditions (with $m = 2.1 \times 10^{-22}$) and the bottom blue curve is under no hoarding conditions (with $m = -7.7 \times 10^{-19}$). The macroeconomic conditions for both curves are almost identical; however, when the tiny differences cause the liquidity stress index to turn positive, bank i's supply curve suddenly shifts from the lower to the upper curve.

Fixing the amount of the overnight loan at $\mathcal{S}^*$ for one bank, Theorem 4 demonstrates that the equilibrium funding rate charged by this bank may jump up discontinuously in response to only minor shocks, such as (a) slight reductions in the reserve available to other dealer banks, (b) an increase in lending by the counterpart bank in wholesale funding markets, or (c) small increments in the per-unit liquidity cost ψ, among other factors influencing the liquidity stress index (see the left panel of Fig. 5). Under the liquidity hoarding condition, even if bank i possesses a substantial initial reserve amount $R_i = \zeta \gg 0$, it will remain conservative in lending out reserves and will charge higher funding rates due to concerns that the early payment a_j by other banks is likely to be low, as illustrated in the right panel of Fig. 5. The next theorem analyzes the general-equilibrium effects on rates and quantities from changes in macroeconomic conditions as reflected through the liquidity stress index.

Theorem 5. *Fix some realization ζ of beginning reserve balances R_i and wholesale borrowing demand $\mathscr{D}$. The equilibrium trading quantity $\mathcal{S}^*(\zeta, \mathscr{D}, m)$ decreases and the funding rate r^* jumps up as a function of the liquidity stress m at the threshold $m = 0$ that triggers liquidity hoarding.*

---

depends on the macroeconomic conditions, $r^*(\zeta, s, m)$ is the funding rate that bank i will charge borrower i for borrowing quantity $S_i = s$ in equilibrium.

More specifically, there exists some $\delta(\zeta, \mathscr{D}) > 0$ such that[23]

$$\lim_{\epsilon^m \downarrow 0} \mathcal{S}^*(\zeta, \mathscr{D}, -\epsilon^m) - \mathcal{S}^*(\zeta, \mathscr{D}, \epsilon^m) > \delta(\zeta, \mathscr{D})$$
$$\lim_{\epsilon^m \downarrow 0} r^*(\zeta, \mathscr{D}, \epsilon^m) - r^*(\zeta, \mathscr{D}, -\epsilon^m) > \xi\delta(\zeta, \mathscr{D}),$$

provided the sets of macroeconomic conditions $\mathcal{M}_{\mathcal{C}}{}^{\epsilon^m}$ are mutually close to each other with respect to the liquidity stress index, and $\zeta - \mathcal{S}^*(\zeta, \mathscr{D}, 0) - Q \neq 0$.

When overall reserve balances are lower, banks not only anticipate receiving lower early payments from their counterpart banks but also start with smaller opening reserves. The equilibrium funding rates are influenced by a combination of two factors: (1) reduced initial reserves of bank i increase its marginal value of liquidity, thereby nonlinearly elevating the funding rates charged by bank i even under the no hoarding condition; (2) diminished reserves of bank j lead bank i to worry about its liquidity management in the payment subgame, causing equilibrium funding rates to spike once the liquidity hoarding condition is triggered. Together, these effects provide a framework for understanding repo rate dynamics throughout 2019.[24]

7 Data

To quantitatively understand the effect of strategic complementarity and calculate the counterfactual sufficient reserves to support effective monetary policy implementation, I estimate my model in the context of the GCF Treasury repo market. This section describes the data I use for this exercise.

I use the GCF repo rate as the proxy for large U.S. dealer banks' lending rate in wholesale funding markets.[25] GCF repo rates data are published by FICC.[26] I obtain two daily trading volume data from the New York Fed: The volumes for calculating the

[23]The equilibrium funding rate function in this theorem is $r^* : \mathbb{R}^3 \to \mathbb{R}^+$, which determines the equilibrium supply curve: given some beginning reserve balances $R_i = \zeta$ and the liquidity stress index m, which depends on the macroeconomic conditions, $r^*(\zeta, \mathscr{D}, m)$ is the equilibrium funding rate at which bank i lends to borrower i, who has a total financing need $D_i = \mathscr{D}$. Similar definitions apply to the equilibrium trading function $\mathcal{S}^* : \mathbb{R}^3 \to \mathbb{R}^+$.

[24]It is worth noting that, according to my data, the demand for repo financing exhibits significant inelasticity. According to my theoretical model, the magnitude of the spikes would be smaller if the demand were more elastic.

[25]The GCF repo market is an interdealer market where large dealers lend to smaller dealers. As a result, the GCF-IOR spread measures the compensation that large dealers require when they provide liquidity by draining reserves.

[26]The FICC-cleared sponsored repo market is a another major inter-dealer market; however, data on

Secured Overnight Financing Rate (SOFR) index and the Triparty General Collateral Rate (TGCR) index. I obtain the quarterly average Treasury repo lending quantity from largest U.S. banks' form 10-Q for a set of large U.S. banks (and large U.S. subsidiaries of foreign banks) including JP Morgan, Bank of America, Goldman Sachs, Morgan Stanley, Citi, Wells Fargo Bank, PNC Bank, Deutsche Bank Trust Company Americas, HSBC Bank USA, and State Street. I combine those two data to estimate the daily repo lending quantity of large dealer banks.

I use two types of information about reserve balances held at the Federal Reserve Banks: daily opening balances and the timing of cash transfers between accounts within each day. Both types of data are provided by Copeland, Duffie and Yang (2020), sourced from the Fedwire Funds Service. Specifically, I observe the total opening reserve balances of the largest 100 accounts managed by depository institutions, and the total opening reserve balances of 10 dealer banks, as identified by Copeland, Duffie and Yang (2020). The total opening reserve balances of 10 dealer banks are held by depository institutions owned by bank holding companies that have a large presence in U.S. repo markets. The total opening balances of these 10 dealer banks is about 40% of total opening balances of the accounts of the 100 largest banks over 2018-19, and the total balances of the 100 accounts are about 85% of total reserves held at Federal Reserve Banks over 2018-19. I call the banks of those 100 accounts other than the 10 dealer banks "other large banks." I do not observe the identities of any banks in my sample.

In addition to daily opening-balance information, Copeland, Duffie and Yang (2020) provide statistics regarding the timing of payments sent over Fedwire within the day. In particular, I observe the time stamp when 50% of the total value of transfers to the 10 dealer banks' accounts has been received by the 10 dealer banks in the day. For example, on September 3, 2019, 50% of the total transfers to the 10 dealer banks had been received by 2:06 pm. I subtract this time stamp from 12:09 pm, the average of this statistic between January 2, 2014, and October 9, 2020, to calculate a measure of payment timing stress, which I call "median time of receives."[27]

I obtain Treasury issuance and redemption data from the Treasury Department. I also

FICC-cleared repo rates are unavailable. Studies such as Senyuz, Anbil and Anderson (2023) provide evidence that sponsored repo rates move closely with the GCF repo rate, particularly in 2019, suggesting that GCF repo rates can serve as a reliable proxy for understanding dynamics within the broader repo market.

[27]This measure is based on standard payment timing metrics used in previous research on intraday payments, such as Armantier, McAndrews and Arnold (2008), McAndrews and Kroeger (2016), and Copeland, Molloy and Tarascina (2019).

obtain total Treasuries outstanding from the U.S. Treasury Fiscal Data and the Treasuries held by the Fed from the Federal Reserve Bank of St. Louis's FRED database. Finally, I obtain total payment volume data from Fedwire.[28] Summary statistics of the key variables are provided in Table 2.

8 Model calibration

The term "calibration" in the section title is selected with consideration. While the parameters of my model are derived using the method of moments—a technique commonly associated with the estimation of model parameters—I am cautious of application of the term "estimation" in the context of this study. This caution stems from the understanding that models featuring endogenous regime shifts, such as the one under consideration, may challenge the direct application of conventional statistical tests, potentially leading to inferences that diverge from standard econometric properties, particularly in finite samples.[29] Despite my meticulous efforts in managing the estimation process, an in-depth study of the econometric properties of the estimators for my model's parameters falls outside this paper's scope. Consequently, the model's quantitative success is assessed mainly based on its out-of-sample prediction performance, in line with approaches commonly found in much of the machine learning literature.

In this study, I focus on a sample spanning from January 3, 2019, to September 18, 2019. I divide this period into two subsamples: the first, from January 3, 2019, to August 31, 2019, which I refer to as the "training dataset," and the second, from September 1, 2019, to September 18, 2019, which I designate as the "test dataset." I estimate the parameters of my model using the method of moments on the training dataset. I chose not to include data from days in previous years within my training dataset, because large dealer banks annually adjust the assumptions underpinning their regulation YY stress testing and resolution planning, which implies that parameter Q in my baseline model may vary from

[28]Without more detailed information, I assume that each of the large banks contributes an equal share of the total payment volume made by the largest 15 banks, estimated at around 68% of the total payment volume over the Fedwire, according to Afonso, Duffie, Rigon and Shin (2022). This implies that each of the large dealer banks in my sample contributes an estimated $1/15 * 68\% = 4.53\%$ of the total payment volume on average each business day.

[29]For example, a common challenge encountered with estimating such models involves a departure from the standard econometric assumption that objective functions exhibit continuity in a small neighborhood of the true parameters. While an examination has verified the continuity of the objective function around my point estimates in this case, there remains a possibility that the bootstrapped standard errors might not fully capture the true variability, and that the finite-sample distribution of the estimators may deviate from normality.

year to year. Moreover, prior to 2019, total reserve balances were considerably higher, diminishing the observable effects of strategic complementarity. To further refine my dataset, I exclude observations from January 1-2, 2019, to avoid the distortive impact of year-end capital requirements, as noted by (Correa, Du and Liao, 2020).[30] I do not include dates after September 18, 2019, in my test dataset, because the Fed reacted to the September repo spikes by directly offering liquidity daily until June 2020.[31] The Fed's reaction distorted the GCF-IOR spread in a way that is outside of my model.

My theoretical framework is capable of characterizing the n-bank model; however, the numerical computation required to find a fixed point in the integral-equation system that characterizes the n-bank equilibrium proves to be time-consuming. Therefore, in this current study, I focus on estimating the two-bank baseline model. As discussed in Appendix D.2, the theoretical outcomes for the n-bank scenario closely align with those derived from the two-bank case.

Given the absence of specific information on bank identities or the distribution of reserves among the largest dealer banks, I proceed with the assumption that the total dealer reserve balances observed daily in my dataset are uniformly distributed across the 10 largest dealer banks. I then categorize these 10 banks into five identical payment pairs, enabling each bank to focus exclusively on the payment strategy of its paired counterpart. Consequently, the model estimation centers on a representative bank pair $\{i, j\}$, under the presumption that they remain unconcerned with the payment strategies of any other large dealer bank $k \notin \{i, j\}$. This simplification significantly reduces computational complexity while effectively preserving the dynamics of strategic complementarity. Specifically, from my sample, on every business day t, I observe the following quantities that serve as inputs to my model for calculating equilibrium:

- Total opening reserve balances R_D^t of 10 dealer banks.
- Total opening reserve balances R_O^t of other large banks.
- Net Treasury issuance T_I^t.
- Total Treasuries outstanding T_D^t.

[30] There are usually acute funding constraints at year-end when GSIBs adjust their balance sheets. This adjustment is due to the scoring that predominantly takes place at year-end to determine the capital surcharge for these institutions in the subsequent years.

[31] The Fed announced it would lend cash to borrowers after most repo trading had occurred on September 17, 2019. Market participants were uncertain about whether the New York Fed would continue its intervention in the following days. Consequently, the GCF repo rates remained elevated until the morning of September 18.

- Total repo lending quantity S^t of the dealer banks.[32]
- Median time of receives $\mathcal{D}_P^t$ of the large dealer banks.
- Month-end date indicator E_{month}^t.

To apply my model to the data, I introduce several additional assumptions. First, inspired by empirical evidence presented by Copeland, Duffie and Yang (2020) on the effects of Treasury issuance and other non-dealer bank reserves, I postulate that each dealer bank i begins business day t with a usable reserve balance calculated as follows:

$$R_i^t = \frac{1}{10}(R_D^t - \underbrace{E_I T_I^t}_{\substack{\text{Impact from} \\ \text{Treasury issuance}}} + \underbrace{E_O R_O^t}_{\substack{\text{Early incoming payment} \\ \text{from large non-dealer banks}}}) - \underbrace{Q}_{\text{normalization}},$$

where E_I, E_O, and Q are parameters to be estimated, representing the payment effects of Treasury issuance, early incoming payments from non-dealer banks, and a normalization constant, respectively. Treasury issuance settlements result in cash transfers from dealer banks' accounts at the Fed to the TGA account, and these transfers must occur near the beginning of the day (Copeland, Duffie and Yang, 2020), significantly draining the reserve balances of large dealer banks. The term $E_I T_I$ captures this effect. In addition, the opening-of-day reserve balances of the other large non-dealer banks are closely linked with early payments received by large dealer banks, evidencing a linear correlation (Copeland, Duffie and Yang 2020, also illustrated in Fig. 16). Large dealer banks can treat the incoming payments from other large banks as part of their usable reserve balances for outgoing payments. The term $E_O R_O$ quantifies the effective augmentation of dealer banks' reserves by early incoming payments from non-dealer banks. The parameter Q largely encapsulates the representative dealer bank's lowest comfortable level of reserves, predominantly influenced by liquidity regulations.[33] However, Q may also incorporate normalization constants related to Treasury issuance and inflows from non-dealer banks. For instance, if the true impact of Treasury issuance can be expressed as $E_I^c + E_I T_I^t$, where E_I^c is a constant, then E_I^c will be effectively subsumed into the value of Q.

Second, I assume that bank i believes bank j's beginning reserve balances follow a normal distribution, $R_j^t \sim \mathcal{N}(R_i^t, \sigma_r)$, and analyze the limiting equilibrium as $\sigma_r \to 0$, in line with the conventions of global game literature (Carlsson and Van Damme, 1993). Consequently, bank i effectively knows the quantity of bank j's reserve balance when day t starts in the limiting equilibrium.[34] Drawing on empirical evidence (Copeland, Duffie

[32]For details on how I calculate S^t, refer to Appendix E.
[33]For details see the Fed's Senior Financial Officer Survey.
[34]I find that altering the value of uncertainty σ_r within a small neighborhood around zero has virtually

and Yang (2020), also illustrated in Table 7) that shows a positive correlation between increased Treasury outstandings (T_D^t) and repo trading volume, I hypothesize that bank i anticipates the repo borrowing demand for borrower j to follow $D_j^t - E_D T_D^t \sim \exp(\lambda)$. This assumption implies that borrowing demand increases linearly with T_D^t and E_D serves as the linear coefficient capturing this relationship.

Third, I use the median time of receives, $\mathcal{D}_P^t$, to infer early payment strategy a_i^*, which is not directly observable. Intuitively, the median time of receives serves as a proxy for measuring the payment delays encountered by large dealer banks on a daily basis, effectively reflecting the early payment behaviors of large banks. Specifically, an increased early payment a_j from bank j to bank i on day t should correspondingly lead to an earlier median time by which bank i receives 50% of its total incoming payments via Fedwire on the same day. To quantify this relationship, I postulate and estimate a linear model for day t as follows:

$$\mathcal{D}_P^t - \mathbb{E}[\mathcal{D}_P^t] = \underbrace{\beta_1^e(a_i^{*t} - \mathbb{E}[a_i^{*t}])}_{\text{early payment from dealer banks}} + \underbrace{\beta_2^e(R_O^t - \mathbb{E}[R_O^t])}_{\text{early payment from non-dealer banks}} + \underbrace{\epsilon_D^t.}_{\text{noise on day } t} \quad . \tag{7}$$

By design, Eq. (7) ensures $\mathbb{E}[\epsilon_D^t] = 0$.

Forth, under the assumption that R_i and D_i possess atomless distributions, the equilibrium funding rates (Eq. (4)) suggest an empirical relationship between GCF repo rates and other observed quantities as follows:

$$(GCF - IOR)^t = \Gamma_i(R_i - Q - S_i^t, \alpha_i^{*t}, \alpha_j^{*t}) + \underbrace{\vartheta^{ME} E_{month}^t}_{\text{month-end effect}} + \underbrace{\vartheta}_{\text{other market factors}} + \underbrace{\epsilon_r^t}_{\text{noise on day } t} . \tag{8}$$

Here, $S_i^t = \frac{1}{10} S^t$ represents the proportionate share of total repo lending attributable to bank i on day t. The term ϑ^{ME} is introduced to capture the month-end seasonality observed in GCF repo rates,[35] while ϑ is estimated to encapsulate the aggregate impact of other market factors influencing GCF repo rates.

Lastly, in lieur of Eq. (4), unobserved repo borrowing demand of borrower i are related

no impact on the estimates and the model's fit.

[35]Before 2020, GCF-IOR spread is usually elevated at month end and quarter end due to regulatory capital requirements on foreign bank holding companies that cause them to reduce their provision of liquidity to interdealer markets, (in particular, a subset of UK banks had their capital ratios calculated based on a snapshot of their month-end balance sheets) as detailed in the studies by Correa, Du and Liao (2020), Ranaldo, Schaffner and Vasios (2020), and Bassi, Behn, Grill and Waibel (2024), among others.

to the equilibrium GCF repo rate and transaction amount as:

$$D_i^t = S_i^t + \frac{(GCF - IOR)^t}{\xi} + \underbrace{\epsilon_{Sm}^t}_{\text{measurement error}}. \tag{9}$$

The structural model has a total of 14 parameters, listed in Table 4. The parameters pertinent to daily payment obligations, λ_N and N_{min}, can be straightforwardly estimated using the method of moments, utilizing payment volume data from Fedwire. The distribution of N_i implies that its mean and variance are $N_{min} + \frac{1}{\lambda_N}$ and $\frac{1}{\lambda_N^2}$, respectively, which are aligned with the corresponding metrics observed in the payment data. Given a candidate set of model parameters, observable inputs from day t, namely $R_D^t, R_O^t, T_I^t, T_D^t, E_{month}^t$, enable me to calculate the equilibrium early payment strategies a_i^{*t}, a_j^{*t} and the corresponding values of Γ_i, Γ_j, as delineated by Theorems 1 and 2. I then incorporate additional observable equilibrium outcome variables, $(GCF - IOR)^t$, S_i^t and $\mathcal{D}_P^t$ to fit two empirical equations Eqs. (7) and (8). This process yields unexplained residuals ϵ_r^t and ϵ_D^t for each day t. My identification assumption is that these residuals, ϵ_r^t and ϵ_D^t, are orthogonal to the above quantities that I use to calculate fit the two empirical equations. This entails setting the finite-sample-analogous expectations of the following moments to zero, effectively leveraging the dataset to ensure the model's predictions correspond with observed behaviors of GCF repo rates and payment system delays:

$$\mathbb{E}\left[\epsilon_r^t\right] = \mathbb{E}\left[\epsilon_r^t R_D^t\right] = \mathbb{E}\left[\epsilon_r^t R_O^t\right] = \mathbb{E}\left[\epsilon_r^t T_I^t\right] = \mathbb{E}\left[\epsilon_r^t E_{month}^t\right] = \mathbb{E}\left[\epsilon_r^t S_i^t\right] = 0$$
$$\mathbb{E}\left[\epsilon_D^t (a_i^{*t} - \mathbb{E}[a_i^{*t}])\right] = \mathbb{E}\left[\epsilon_D^t ({R_O}^t - \mathbb{E}[{R_O}^t])\right] = \mathbb{E}\left[\epsilon_D^t R_D^t\right] = \mathbb{E}\left[\epsilon_D^t E_{month}^t\right] = 0$$

I further assume that the measurement error $\epsilon_{S_m}^t$, as defined in Eq. (9), has a zero mean and is uncorrelated with the outstanding Treasury debt T_D^t. These assumptions introduce two additional moments that are crucial for identifying the parameters λ and E_D.[36]

$$\mathbb{E}\left[S_i^t + (GCF - IOR)^t/\xi - E_D T_D^t - \lambda^{-1}\right] = \mathbb{E}\left[(S_i^t + (GCF - IOR)^t/\xi - E_D T_D^t - \lambda^{-1}) T_D^t\right] = 0$$

I estimate the model based on the test data set of my sample (business days from January 3, 2019, to August 31, 2019). The point estimates are recorded in Table 4. Admittedly, the GCF repo market is a complicated OTC market, featuring relationship trading (Paddrik, Ramirez, McCormick et al., 2021), search frictions (Afonso and Lagos, 2015),

[36] Based on the assumptions regarding the probability distribution of the borrowing demands D_i^t and D_j^t, it follows from Eq. (9) that $\mathbb{E}[-\epsilon_{S_m}^t] = \mathbb{E}[S_i^t + (GCF - IOR)^t/\xi] - \mathbb{E}[D_i^t]$. Since $\epsilon_{S_m}^t$ and demeaned D_i^t are both uncorrelated with T_D^t, $\mathbb{E}[(-\epsilon_{S_m}^t + D_i^t - E_D T_D^t - \lambda^{-1}) T_D^t] = \mathbb{E}[(S_i^t + (GCF - IOR)^t/\xi - E_D T_D^t - \lambda^{-1}) T_D^t] = 0$.

and market segmentation (Han, 2020; Avalos, Ehlers and Eren, 2019; Duffie and Krishnamurthy, 2016). My model abstracts away from those frictions to focus only on the relationship between the quantity of reserve supply and GCF-IOR spread. However, the goal of my quantity exercise is not to provide the most accurate quantitative model to describe the GCF repo market. Rather, my focus is on exploring the impact of strategic complementarity on overnight wholesale funding rate spreads. Nevertheless, my model fits the in-sample variations of GCF-IOR spread and median payment timing simultaneously reasonably well on my test data set. Table 5 compares the goodness of fit of my model with three other types of models: (1) one linear model for the GCF-IOR spread and one linear model for the median time of receives, (2) a machine learning model of random forest trained using cross-validation, and (3) an otherwise identical model of mine shutting down the effect of strategic complementarity (see details of this model in Appendix D.3). The measures of goodness of fit include mean squared error, correlation between the model-predicted and actual GCF-IOR spread, and correlation between the model-predicted and actual median time of receives. For linear models, I calculate R^2.[37] My model is the only one that is able to simultaneously fit both the GCF-IOR spread and the median time of receives, while achieving similar performance relative to other models in each dimension.

Although I excluded the September repo spike events from my sample when estimating my model, my quantitative model correctly predicts the repo spikes on September 16-19, 2019, as an out-of-sample event. Figs. 6 and 7 compare the performance across all four models. In these two plots, the days to the left of the dashed vertical line are my training dataset, and the days to the right are my test dataset. To study the out-of-sample performance of all aforementioned models, I use the out-of-sample actual observable set $\{R_D^t, R_O^t, T_I^t, T_D^t, E_{month}^t, S_i^t\}$ from my test dataset as inputs into my model and three other models with parameters estimated from the training dataset. It is important to note that although the Fed does not observe these inputs on day $t-1$, predicting these variables is straightforward because they are either prescheduled or exhibit stable time series characteristics. The predicted GCF-IOR spread generated by all four models tracks the variation of the actual GCF-IOR spread closely in sample. However, only my model successfully predicts significant spikes in the GCF-IOR spread from September 16 to September 19, 2019.[38]

[37] For completeness, I also calculate Pseudo-R^2 defined by Schabenberger and Pierce (2001) for nonlinear models, although Pseudo-R^2 is not a great measure for comparing different nonlinear models.

[38] My model reasonably captures the GCF-IOR spikes on September 16 and September 19, 2019, with prediction errors of approximately 20 basis points for each instance. However, it fails to replicate the

The performance of my quantitative model on the test dataset validates the mechanism of my theory. Without prior knowledge of how strategic complementarity in interbank intraday payment timing works, simple statistical models and even machine learning models are unlikely to capture the nonlinear or discontinuous reactions of short-term wholesale funding rates when reserves balances are close to being insufficient. Thus, those models tend to *underestimate* the level of sufficient reserves. Because the Fed has access to granular data of intraday interbank payment timing, real-time changes of every bank's reserve account balance, and activities in the short-term funding markets, a good understanding of strategic complementarity in interbank intraday payment timing could have allowed the Fed to develop a full-fledged quantitative model to shape its policy of reserve supply. An effective quantitative model of this nature, which helps predict funding rate distortions out of sample, serves as a valuable tool for central bankers, facilitating precautionary interventions.

9 What quantity of reserves would have been sufficient?

Copeland, Duffie and Yang (2020) note the reserve balances of the top 10 repo active dealer banks are more important in directly determining the spreads of various repo rates over IOR than reserve balances of other non-dealer large banks. My quantitative model enables the estimation of the minimum reserve levels necessary for large dealer banks to maintain short-term funding rates close to the Fed's policy target and to support an efficient payment system. This level may vary daily due to changes in macroeconomic conditions, such as fluctuations in borrowing demands within the wholesale funding markets, variations in Treasury issuance, and adjustments in the reserve balances of other non-dealer large banks.

Fig. 8 shows the sufficient amount of reserves of large dealer banks required to keep the expected GCF-IOR spread below 13 basis points, 26 basis points. and 52 basis points,

substantial spike on September 17—the actual spike was 390.7 basis points, but the model predicts only 91.8 basis points. This discrepancy is likely attributable to several factors: (1) The model assumes linear delay and regulatory costs with coefficients c and ψ, whereas the actual cost function may be nonlinear. (2) Due to limitations in my dataset, I have assumed that reserves are evenly distributed among the large dealer banks; however, extra reserves, i.e., reserves beyond regulatory requirements, may became concentrated in a few banks in 2019, as indicated by banks' quarterly reports. This uneven distribution may have exacerbated liquidity hoarding and driven up funding rates. (Interested readers can contact the author for some theoretical results on reserve concentration effects.) (3) The baseline model abstracts away certain market frictions, such as market power, which might have been particularly influential on September 17, 2019, when the concentration of extra reserves meant that only a few banks held sufficient liquidity.

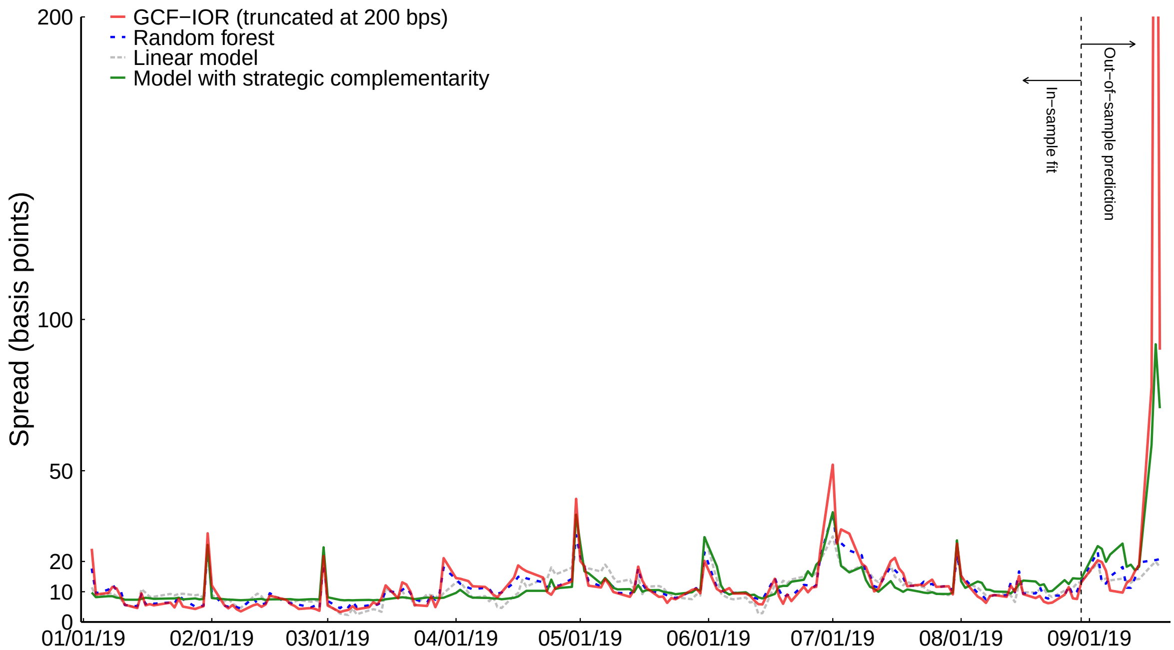

Figure 6: Comparison of in-sample and out-of-sample expected GCF-IOR spreads conditional on $\{R_D^t, R_O^t, T_I^t, T_D^t, S_i^t, E_{month}^t\}$ generated by three models: (1) my model incorporating strategic complementarity, (2) a linear regression model, and (3) a random forest machine learning model. The delineation between in-sample (to the left) and out-of-sample (to the right) periods is marked by a dashed line, representing August 31, 2019.

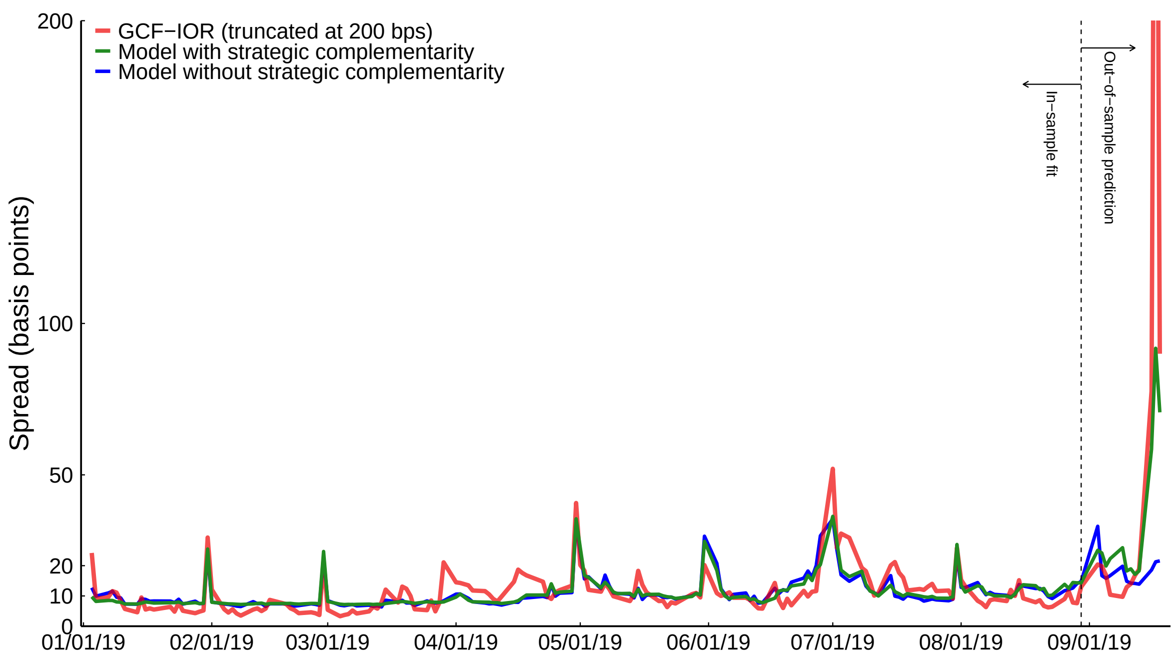

Figure 7: The fitted expected GCF-IOR spreads (conditional on $\{R_D^t, R_O^t, T_I^t, T_D^t, S_i^t, E_{month}^t\}$) from my model with strategic complementarity and an otherwise identical model without strategic complementarity. In-sample days are to the left of the dashed line; out-of-sample days are the days to the right. Dashed line: August 31, 2019.

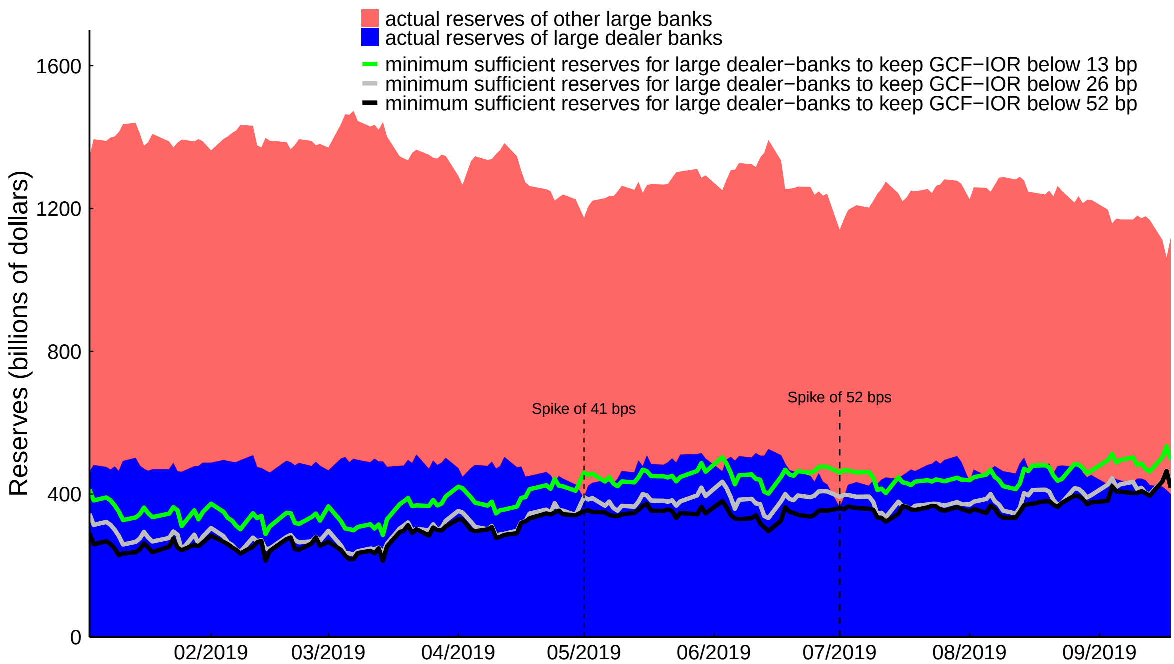

Figure 8: Counterfactual minimum levels of reserves of large dealer banks required to maintain the expected GCF-IOR spread below 13 bps, 26 bps, and 52 bps, respectively, holding other variables $(\{R_O^t, T_I^t, T_D^t, S_i^t, E_{month}^t\})$ at their actual historical levels. This analysis intentionally excludes month-end days to focus solely on the influences of reserve supply and payment delays.

respectively, as predicted by my model.[39] My estimated model suggests that beginning in September 2019, the historical reserve balances of the large dealer banks became insufficient to keep the expected GCF-IOR spread below 13 basis points, as confirmed by the data. All three lines in the plot increase progressively, because, up to September 18, 2019, reserve balances of other large banks declined substantially, while repo borrowing demand steadily increased throughout 2019. Treasury issuance also temporarily elevates the level of necessary reserves.

Note that on July 1, 2019, dealer banks' total reserve balances were lower than those of September 17, 2019, but the GCF-IOR spread on July 1 was much lower than the spread on September 17. This observation would be a puzzle through the lenses of models that study the repo market in isolation. My model provides an explanation for this observation: The reserve balances of other large banks were larger on July 1, so large dealer banks expected they could rely on early incoming payments from other large banks to make their outgoing payments. Therefore, large dealer banks did not hoard reserves in paying other banks on July 1, and the GCF-IOR spread did not spike with a large magnitude (though with

[39]Due to noise that is not captured by my theoretical model, the actual GCF-IOR spread may fluctuate above and below the expected GCF-IOR spread.

a smaller amount of opening reserve balances, the dealer banks demanded higher repo rates). On September 16, 2019, however, a combined effect of low reserve balances of all large banks and increased repo borrowing demand triggered the liquidity hoarding condition, consistent with my model's prediction, thus causing the large repo spike on that day. By integrating the interbank payment market and wholesale funding markets, my model captures both events. Fig. 8 shows the sufficient level of reserves for large dealer banks is higher on September 16-18 than on July 1.

10 Conclusion

The post-GFC liquidity rules and supervision significantly increase the incentives of large U.S. dealer banks to maintain substantial intraday reserve buffers. I show that a sufficiently low supply of reserves causes banks to suddenly hoard reserves, reinforced by a feedback effect stemming from the strategic complementarity of intraday payment timing, and leads to intraday payment timing stress.

My main results suggest that to avoid reserve hoarding and wholesale funding rate spikes, the Fed would want to ensure banks have enough reserves to meet (1) intraday interbank payment needs and (2) borrowing demand in wholesale funding markets (Anbil, Anderson, Cohen and Ruprecht, 2023). I show that factors determining reserve sufficiency can be summarized by one liquidity stress index. To reduce frictions in monetary policy implementation and the interbank payment system, the Fed may also relax post-crisis liquidity regulations to encourage the use of the Fed's intraday overdraft facility and reduce large banks' dependence on incoming payments in sending their outgoing payments.

In July 2021, the Federal Reserve established the Standing Repo Facility (SRF) as a backstop in the repo markets. However, borrowing from the SRF could expand the balance sheets of large dealer banks, potentially exacerbating their capital constraints by, for example, increasing their supplementary leverage ratio. There is also concern over a potential "stigma" associated with borrowing from the Fed. Given the uncertainties regarding the SRF's effectiveness during financial stress and the Fed's intention not to intervene regularly in funding markets, it remains crucial for the Fed to ensure an adequate reserve supply.

My paper is closely related to the current dialogue concerning the Federal Reserve's quantitative tightening strategy. During the March 2024 FOMC press conference, Federal Reserve Chair Jerome Powell emphasized the goal to conclude QT once reserves reach "the lowest possible ample number." In light of this, my paper contributes a quantitative framework to assess reserve ampleness with the objectives of ensuring an efficient interbank payment system, effective monetary policy implementation, and liquidity in wholesale funding markets.

References

Acharya, Viral V and Ouarda Merrouche, "Precautionary hoarding of liquidity and interbank markets: Evidence from the subprime crisis," *Review of finance*, 2013, *17* (1), 107–160.

__ **, Rahul Singh Chauhan, Raghuram G Rajan, and Sascha Steffen**, "Liquidity dependence: Why shrinking central bank balance sheets is an uphill task," *Available at SSRN*, 2022.

Adrian, Tobias and Hyun Song Shin, "Financial intermediary balance sheet management ," *Annu. Rev. Financ. Econ.*, 2011, *3* (1), 289–307.

Afonso, Gara and Hyun Song Shin, "Precautionary Demand and Liquidity in Payment Systems," *Journal of Money, Credit and Banking*, 2011, *43*, 589–619.

__ **and Ricardo Lagos**, "Trade Dynamics in the Market for Federal Funds," *Econometrica*, 2015, *83*, 263–313.

__ **, Darrell Duffie, Lorenzo Rigon, and Hyun Song Shin**, "How abundant are reserves? Evidence from the wholesale payment system," Technical Report, National Bureau of Economic Research 2022.

__ **, Kyungmin Kim, Antoine Martin, Ed Nosal, Simon Potter, and Sam Schulhofer-Wohl**, "Monetary Policy Implementation with an Ample Supply of Reserves," Staff Report 910, Federal Reserve Bank of New York January 2020.

__ **, Marco Cipriani, Adam Copeland, Anna Kovner, Gabriele La Spada, and Antoine Martin**, "The Market Events of Mid-September 2019," FRB of New York Staff Report 918, Federal Reserve Bank of New York March, 2020.

Anbil, Sriya, Alyssa Anderson, and Zeynep Senyuz, "Are Repo Markets Fragile? Evidence from September 2019," Working Paper, Federal Reserve Board, October 2020.

__ **, __ , and __** , "What Happened in Money Markets in September 2019?," FEDS Notes, Board of Governors of the Federal Reserve System, February, 2020.

__ **, __ , Ethan Cohen, and Romina Ruprecht**, "Stop Believing in Reserves," Technical Report, Working paper 2023.

Andros, Joseph, Michael Beall, Francis Martinez, Tony Rodrigues, Mary-Frances Styczynski, and Alex Thorp, "Approaches to Estimating Aggregate Demand for Reserve Balances," FEDS Notes, October 17, 2019.

Angeletos, George-Marios and Iván Werning, "Crises and prices: Information aggregation, multiplicity, and volatility," *american economic review*, 2006, *96* (5), 1720–1736.

— , **Christian Hellwig, and Alessandro Pavan**, "Signaling in a global game: Coordination and policy traps," *Journal of Political economy*, 2006, *114* (3), 452–484.

Armantier, Olivier, James McAndrews, and Jeffrey Arnold, "Changes in the Timing Distribution of Fedwire Funds Transfers," *Economic Policy Review*, 2008, *14* (2).

Ashcraft, Adam, James McAndrews, and David Skeie, "Precautionary Reserves and the Interbank Market," *Journal of Money, Credit and Banking*, 2011, *43*, 311–348.

Atalay, Enghin, Antoine Martin, and James McAndrews, "Quantifying the Benefits of a Liquidity-Saving mechanism ," *FRB of New York Staff Report*, 2010.

Avalos, Fernando, Torsten Ehlers, and Egemen Eren, "September stress in dollar repo markets: passing or structural? ," *BIS Quarterly Review*, December 8, 2019.

Baklanova, Viktoria, Adam Copeland, and Rebecca McCaughrin, "Reference Guide to US Repo and Securities Lending Markets," Staff Report 740, September, 2015.

Bassi, Claudio, Markus Behn, Michael Grill, and Martin Waibel, "Window dressing of regulatory metrics: evidence from repo markets," *Journal of Financial Intermediation*, 2024, p. 101086.

Bech, Morten, "Intraday Liquidity Management: A Tale of Games Banks Play," *Economic Policy Review*, 2008, *14.*

— **and Rod Garratt**, "The Intraday Liquidity Management Game," *Journal of Economic Theory*, 2003, *109*, 198–219.

Bech, Morten L, Antoine Martin, and James McAndrews, "Settlement Liquidity and Monetary Policy Implementation—Lessons from the Financial Crisis," *Economic Policy Review*, 2012, *18* (1).

Board of Governors of the Federal Reseve System, "History of the FOMC's Policy Normalization Discussions and Communications," Federal Reserve Board, Washington, D.C., October 2019.

Börgers, Tilman, *An introduction to the theory of mechanism design*, Oxford University Press, USA, 2015.

Bush, Ryan, Adam Kirk, Antoine Martin, Philip Weed, and Patricia Zobel, "Stressed Outflows and the Supply of Central Bank Reserves," *Liberty Street Economics*, February 20, 2019.

Carlsson, Hans and Eric Van Damme, "Global games and equilibrium selection," *Econometrica: Journal of the Econometric Society*, 1993, pp. 989–1018.

Carpenter, Seth B, Selva Demiralp, and Zeynep Senyuz, "Volatility in the federal funds market and money market spreads during the financial crisis," *Journal of Financial Stability*, 2016, *25*, 225–233.

Copeland, Adam, Darrell Duffie, and Yilin (David) Yang, "Reserves were not so ample after all," Working Paper, Stanford Graduate School of Business in preparation, 2020.

— , **Linsey Molloy, and Anya Tarascina**, "What Can We Learn From the Timing of Interbank Payments?," *Liberty Street Economics*, Federal Reserve Bank of New York February, 2019.

Correa, Ricardo, Weixin Du, and Gordon Liao, "US Banks and Global Liquidity," Working Paper, Federal Reserve Board and University of Chicago, March, 2020.

d'Avernas, Adrien and Quentin Vandeweyer, "Intraday Liquidity and Money Market Dislocations ," Working Paper November 2020.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan, "Deviations from covered interest rate parity," *The Journal of Finance*, 2018, *73* (3), 915–957.

Duffie, Darrell, "Special repo rates ," *The Journal of Finance*, 1996, *51* (2), 493–526.

__ , "Post-Crisis Bank Regulations and Financial Market Liquidity," Baffi Lecture, Banca d'Italia, March, 2018.

__ **and Arvind Krishnamurthy**, "Passthrough Efficiency in the Fed's New Monetary Policy Setting," in Richard A. Babson, ed., *Designing Resilient Monetary Policy Frameworks for the Future, A Symposium Sponsored by the Federal Reserve Bank of Kansas City*, Jackson Hole, Wyoming, August 25-27, 2016, Federal Reserve Bank of Kansas City, August 2016, pp. 21–102.

Egelhof, James, Antoine Martin, and Noah Zinsmeister, "Regulatory Incentives and Quarter-End Dynamics in the Repo Market," Federal Reserve Bank of New York, *Liberty Street Economics*, Federal Reserve Bank of New York, August 2017.

Goldstein, Itay and Ady Pauzner, "Demand–Deposit Contracts and the Probability of Bank Runs," *the Journal of Finance*, 2005, *60* (3), 1293–1327.

__ **, Ming Yang, and Yao Zeng**, "Payments, reserves, and financial fragility," *Available at SSRN 4547329*, 2023.

Gorton, Gary and Andrew Metrick, "Securitized banking and the run on repo," *Journal of Financial Economics*, 2012, *104*, 425–451.

Han, Yangjue, "Monetary Policy Transmission Before and After the Crisis," Working Paper, New York University, May 2020.

He, Zhiguo and Wei Xiong, "Dynamic debt runs," *The Review of Financial Studies*, 2012, *25* (6), 1799–1843.

__ **, Arvind Krishnamurthy, and Konstantin Milbradt**, "A model of safe asset determination," *American Economic Review*, 2019, *109* (4), 1230–62.

Heider, Florian, Marie Hoerova, and Cornelia Holthausen, "Liquidity hoarding and interbank market rates: The role of counterparty risk," *Journal of Financial Economics*, 2015, *118* (2), 336–354.

Ihrig, Jane, "Banks' Demand for Reserves in the Face of Liquidity Regulations," *On the Economy Blog*, Federal Reserve Bank of St. Louis, March 5, 2019.

__ **, Zeynep Senyuz, and Gretchen Weinbach**, "The Fed's "Ample-Reserves" Approach to Implementing Monetary Policy," Technical Report, Board of Governors of the Federal Reserve System, February 2020.

Ivashina, Victoria, David S Scharfstein, and Jeremy C Stein, "Dollar funding and the lending behavior of global banks ," *The Quarterly Journal of Economics*, 2015, *130* (3), 1241–1281.

Kahn, R Jay and Luke Olson, "Who Participates in Cleared Repo?," Technical Report, Office of Financial Research, US Department of the Treasury 2021.

Lenel, Moritz, Monika Piazzesi, and Martin Schneider, "The short rate disconnect in a monetary economy," *Journal of Monetary Economics*, 2019, *106*, 59–77.

Liu, Xuewen, "Interbank market freezes and creditor runs," *The Review of financial studies*, 2016, *29* (7), 1860–1910.

Logan, Lorie, "Observations on Implementing Monetary Policy in an Ample-Reserves Regime," Remarks before the Money Marketeers of New York University, Federal Reserve Bank of New York, April 2019.

Lopez-Salido, David and Annette Vissing-Jorgensen, "Reserve demand, interest rate control, and quantitative tightening," *Interest Rate Control, and Quantitative Tightening (February 27, 2023)*, 2023.

Martin, Antoine, McAndrews James, Ali Palida, and David Skeie, "Explaining the Puzzling Behavior of Short-Term Money Market Rates," *Liberty Street Economics*, Federal Reserve Bank of New York, August 24, 2020.

McAndrews, James and Alexander Kroeger, "The payment system benefits of high reserve balances," *Journal of Payments Strategy & Systems*, 2016, *10* (1), 72–83.

_ **and Samira Rajan**, "The timing and funding of Fedwire funds transfers ," *Economic Policy Review*, 2000, *6* (2).

_ **and Simon Potter**, "Liquidity Effects of the Events of September 11, 2001," *Economic Policy Review*, 2002, *8.*

Milgrom, Paul and Ilya Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, 2002, *70*, 583–601.

Morris, Stephen and Hyun Song Shin, "Unique equilibrium in a model of self-fulfilling currency attacks," *American Economic Review*, 1998, pp. 587–597.

_ **and** _ , "Global games: Theory and applications," 2001.

Mussa, Michael and Sherwin Rosen, "Monopoly and product quality," *Journal of Economic theory*, 1978, *18*, 301–317.

Myerson, Roger B, "Multistage games with communication," *Econometrica*, 1986, *54*, 323–358.

Nelson, Bill and Francisco Covas, "Bank Regulations and Turmoil in Repo Markets ," Bank Policy Institute Blog 2019. September 26, 2020.

Nicolae, Laura, "The Effects of Liquidity Regulation on Bank Demand for Reserves and Federal Reserve Balance Sheet Policy," Working Paper, May 2020.

Paddrik, Mark, Carlos Ramirez, Matthew McCormick et al., "The Dynamics of the US Overnight Triparty Repo Market," *FEDS Notes*, 2021, p. 02.

Piazzesi, Monika, Ciaran Rogers, and Martin Schneider, "Money and banking in a New Keynesian model," 2019.

Pozsar, Zoltan, "Collateral Supply and o/n Rates," Global Money Notes 22, Credit Suisse Economics, May 2019.

— , "Design Options for an o/n Repo Facility," Global Money Notes Number 25, Credit Suisse Economics, September 2019.

Ranaldo, Angelo, Patrick Schaffner, and Michalis Vasios, "Regulatory Effects on Short-Term Interest Rates," *Journal of Financial Economics*, 2020, *forthcoming.*

Schabenberger, Oliver and Francis J Pierce, *Contemporary statistical models for the plant and soil sciences*, CRC press, 2001.

Schoenmaker, Dirk, "A comparison of alternative interbank settlement systems," *FMG DP*, 1995, *204.*

Senyuz, Zeynep, Sriya Anbil, and Alyssa G Anderson, "The Importance of Repo Trading Relationships During Market Stress," *Available at SSRN 4572075*, 2023.

Soramäki, Kimmo, Morten L Bech, Jeffrey Arnold, Robert J Glass, and Walter E Beyeler, "The topology of interbank payment flows," *Physica A: Statistical Mechanics and its Applications*, 2007, *379* (1), 317–333.

Wallen, Jonathan, "Markups to Financial Intermediation in Foreign Exchange Markets," Working Paper, Stanford Graduate School of Business, January 2020.

Younger, Joshua, Henry St. John, and Sejal Aggarwal, "If it's not one thing, it's another. The payment system risks of a repo shock," J.P. Morgan North America Fixed Income Strategy Report 2020. July, 14, 2020.

Zhou, Ruilin, "Understanding intraday credit in large-value payment systems," *Federal Reserve Bank of Chicago Economic Perspectives*, 2000, *24* (3), 29–44.

Appendix: Table of contents

A Appendix: Tables and figures

Table 1: Summary statistics for sample used by Copeland, Duffie and Yang (2020)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
dealer opening balances ($ billions)	2,046	819.3	251.3	362.0	664.4	744.4	1,051.3	1,378.0
other non-dealer large bank balances ($ billions)	2,046	1,266.6	277.5	652.4	1,067.3	1,265.9	1,477.9	1,836.1
Tbills outstanding ($ billions)	2,059	2,711.9	1,166.9	1,233.0	1,742.0	2,274.0	3,810.7	4,984.4
Bill issuance ($ billions)	2,059	40.1	61.7	−0.03	−0.001	0.0	92.0	272.9
Coupon issuance ($ billions)	2,059	12.6	40.4	−0.02	0.0	0.0	0.0	301.3
Treasuries redemption ($ billions)	2,059	47.1	65.6	0.0	0.0	0.0	101.5	342.6
median time of receives (minutes)	2,046	0.7	51.1	−107.4	−36.4	−4.4	39.6	154.6
SOFR - IOR (basis points)	2,061	−8.3	11.7	−29	−15	−10	−2	315
GCF - IOR (basis points)	2,059	−3.1	15.1	−47.4	−10.2	−5.4	2.7	390.7
Treasuries issuance ($ billions)	2,059	52.7	73.2	−0.03	0.0	0.0	107.6	470.1
quarter-end fixed effect	2,060	0.02	0.1	0	0	0	0	1
corporate tax to US treasury ($ billions)	2,060	1.3	5.5	−0.1	0.1	0.1	0.4	63.7
dealer bank deposits uninsured ($ billions)	2,061	2,214.0	478.2	1,685.8	1,828.7	1,933.3	2,773.1	3,039.9

Note: This table includes days from January 1, 2015, to March 31, 2023.

Table 2: Summary statistics for the sample used in Section 8

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
dealer opening balances ($ billions)	179	469.9	27.7	362.0	452.0	475.7	489.8	525.4
other non-dealer large bank balances ($ billions)	179	831.4	66.5	652.4	782.5	810.8	890.4	975.2
Total Treasury outstanding ($ billions)	179	14,112.2	192.2	13,852.5	14,026.8	14,066.7	14120.4	14,621.1
Bill issuance ($ billions)	179	35.8	42.5	−0.01	−0.001	0.0	79.0	117.0
Coupon issuance ($ billions)	179	11.3	34.4	−0.003	0.0	0.0	0.0	156.9
Treasuries redemption ($ billions)	179	43.6	51.3	0.0	0.0	0.0	84.0	245.7
median time of receives (minutes)	179	74.5	25.4	−2.4	58.6	72.6	90.6	150.6
SOFR - IOR (basis points)	179	7.0	24.2	−3	1	4	7	315
GCF - IOR (basis points)	179	14.1	30.1	3.3	6.6	9.9	13.4	390.7
Treasuries issuance ($ billions)	179	47.1	55.1	−0.01	−0.001	25.0	84.5	253.0
quarter-end fixed effect	179	0.01	0.1	0	0	0	0	1
corporate tax to US treasury ($ billions)	179	1.1	4.2	0.02	0.05	0.1	0.3	34.7
dealer bank deposits uninsured ($ billions)	179	1,934.0	18.0	1,909.0	1,918.8	1,929.2	1,948.8	1,973.8

Note: This table includes days from January 3, 2019, to September 18, 2019.

Table 3: Quantile regression for the 99th percentile of repo spreads

Dependent variable:	SOFR - IOR			GCF - IOR		
	(1)	(2)	(3)	(4)	(5)	(6)
dealer opening balances	-34.2^{***}	-77.3^{***}	-41.8^{***}	-49.4^{***}	-98.9^{***}	-46.6
	(4.04)	(12.4)	(12.4)	(10.1)	(30.5)	(28.9)
quarter-end fixed effect	39.8^{**}	29.2	37.1^{***}	240.0^{***}	226.0^{**}	240.0^{***}
	(17.4)	(22.1)	(5.1)	(48.5)	(90.5)	(14.8)
median time of receives			0.129^{***}			0.179^{**}
			(0.0415)			(0.0742)
Tbills outstanding		3.01	-0.645		1.86	-2.63
		(2.65)	(2.98)		(6.29)	(6.45)
Treasuries redemption		-25.1	1.52		-84.7	-73.2
		(36.7)	(61.8)		(107.0)	(102.0)
Bill issuance		30.6	11.4		84.2	71.6
		(33.0)	(55.4)		(96.5)	(92.6)
Coupon issuance		52.4	18.0		103.0	71.3
		(35.6)	(51.6)		(116.0)	(101.0)
dealer bank deposits uninsured		26.5^{***}	13.4^{*}		31.6	12.1
		(8.54)	(7.86)		(20.1)	(19.6)
corporate tax to US treasury		-34.3	-12.4		716.0	828.0
		(209.0)	(189.0)		(1939.0)	(1412.0)
Constant	36.6^{***}	4.76	13.8^{*}	58.3^{***}	24.5	34.6^{**}
	(4.88)	(6.59)	(7.83)	(12.5)	(16.2)	(17.1)
Observations	2046	2043	2039	2045	2042	2038
pseudo-R^1	0.289	0.345	0.384	0.016	0.0811	0.141

SOFR is the secured overnight financing rate and IOR is interest on reserves. SOFR-IOR and GCF-IOR are in basis points. The units of the explanatory variables are trillions of dollars and minutes. The left panel is from Copeland, Duffie and Yang (2020). The right panel replicates the results using the GCF-IOR spread. The sample is from 01/01/2015 to 03/31/2023.

Table 4: Calibrated parameters for the model detailed in Section 8

Parameters	Meaning	Point Estimate	Bootstrap Standard Error
ξ	governs demand elasticity	319.0 (basis point)	13.6
c	late payment cost	278.4 (basis point)	10.1
Q	regulatory minimum	16.2 (billion dollar)	4.74
$1/\lambda$	$\mathbb{E}[D_i - D_{min}]$	21.9 (billion dollar)	1.68
ϑ^{ME}	month-end effect	17.1 (basis point)	1.95
ψ	regulatory cost	$21325.1	180818.7
E_I	Treasury issuance effect	0.024	0.169
E_O	early payment from other banks is $E_O R_O$	0.71	0.061
E_D	$D_{min} = E_D T_D$	0.0317	0.00169
β_1^e	coefficient	-32.5	6.94
β_2^e	coefficient	-62.80	8.00
ϑ	other factors	6.90 (basis point)	10.1
N_{min}	minimum total payment volume	29.1 ($bn)	1.45
$1/\lambda_N$	$\mathbb{E}[N_i - N_{min}]$	8.4 (billion dollars)	1.47

Note: When configuring inputs for my model, parameter values—including Q, λ, ψ, N_{min}, and λ_N—should be adjusted to reflect amounts in trillions of dollars (e.g., input Q as 0.0162). This adjustment ensures consistency with the model's input variables $\{R_D^t, R_O^t, T_I^t, T_D^t, S_i^t\}$, which are also specified in trillions of dollars during execution.

For a clearer economic interpretation, the parameter ψ is listed in the table as the dollar penalty per unit of overdraft, equating to a substantial value in basis points per trillion dollars. To determine the accuracy of parameter estimators, I employ a bootstrap methodology, executed with 2,000 bootstrap samples. The bootstrap distribution of ψ is characterized by right skewness and a pronounced fat tail, contributing to a notably large bootstrap standard error.

Regarding the predictive accuracy of the model, approximately 9.2% of the bootstrapped parameter values predict a spike in repo rates on September 11, 2019, 7.5% on September 12, 2019, 35.6% on September 13, 2019, 94.9% on September 16, 2019, 96.0% on September 17, 2019, and 94.7% on September 18, 2019. These results suggest a non-trivial risk of a Type I error (false-positive prediction of a large spike) for September 13, 2019, according to the bootstrap statistics, with a minimal likelihood of Type II errors (false-negative predictions) during September 16-19, 2019.

Table 5: In-sample model fit and out-of-sample performance for four types of models: linear models, my model with strategic complementarity, a model without strategic complementarity, and a random forest machine learning model.

Panel A: In-sample fit

	MSE for fitted GCF-IOR	R^2 for fitted GCF-IOR	Correlation between fitted and actual GCF-IOR	MSE for fitted median time of receives	R^2 for fitted median time of receives	Correlation between fitted and actual median time of receives
model with strategic complementarity	17.86	0.60	0.78	0.64	0.36	0.60
model without strategic complementarity	19.10	0.58	0.76			
linear model for GCF-IOR	22.09	0.62	0.72			
linear model for median time of receives				0.59	0.42	0.65
random forest	18.47	0.59	0.96			

Note: R^2 for nonlinear models is the Pseudo-R^2 defined by Schabenberger and Pierce (2002).
MSE stands for mean squared error. For random forest, MSE and R^2 are based on the out-of-bag prediction error.
The linear models used here correspond to the second and forth columns from Table 6 respectively.
The data for this panel is from 01/03/2019 to 08/31/2019.

Panel B: Out-of-sample mean squared error (MSE)

	model with strategic complementarity	model without strategic complementarity	linear model for GCF-IOR	random forest
MSE	7548.63	12082.31	12185.35	12110.79

Note: Units are squared basis points for each column.
The data for this panel is from 09/01/2019 to 09/18/2019.

Table 6: Linear models for GCF-IOR and payment delays

	GCF-IOR		median time of receives	
	(1)	(2)	(3)	(4)
dealer opening balances	-124.0^{***}	-118.0^{***}	-1.75	-2.00
	(28.3)	(27.2)	(2.15)	(2.18)
large non-dealer bank balances	-29.5^{***}	-47.1^{***}	-9.27^{***}	-7.76^{***}
	(4.38)	(6.29)	(1.12)	(1.58)
net Treasury issuance	68.4^{**}	94.1^{***}	-9.64^{**}	-11.4^{**}
	(32.2)	(33.3)	(4.33)	(4.50)
Treasuries outstanding		-11.7^{***}		0.655
		(2.48)		(0.637)
repo lending quantity		8.96		2.88
		(12.1)		(2.45)
month-end fixed effect	15.7^{***}	14.5^{***}	0.297	0.273
	(2.16)	(2.27)	(0.349)	(0.346)
constant	93.5^{***}	267.0^{***}	11.6^{***}	0.316
	(13.4)	(44.5)	(1.02)	(9.87)
Observations	166	166	166	166
R^2	0.58	0.618	0.376	0.386
Adjusted R^2	0.569	0.603	0.361	0.363
Residual Std. Error	4.41	4.23	0.80	0.798

Note: Standard errors are adjusted for heteroskedasticity. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.
Units are trillions of dollars and minutes. Sample period: 01/03/2019-08/31/2019.

Table 7: Basic regression models for repo trading volume.

	Volume		
	(1)	(2)	(3)
Treasuries Outstanding	0.0799^{***}	0.0312^{***}	0.0295^{***}
	(0.00633)	(0.0102)	(0.0102)
Large Banks' Reserve Balances		-0.186^{***}	-0.183^{***}
		(0.0372)	(0.0385)
Treasury Issuance			0.108
			(0.173)
Treasury Redemptions			-0.00915
			(0.193)
Observations	153	153	153
R^2	0.571	0.623	0.626
Adjusted R^2	0.568	0.618	0.616
Residual Std. Error	62.8	59.1	59.2

Note: Standard errors are adjusted for heteroskedasticity. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.
Constant included for each specification. Sample: biweekly 09-28-2016 to 09-18-2019.

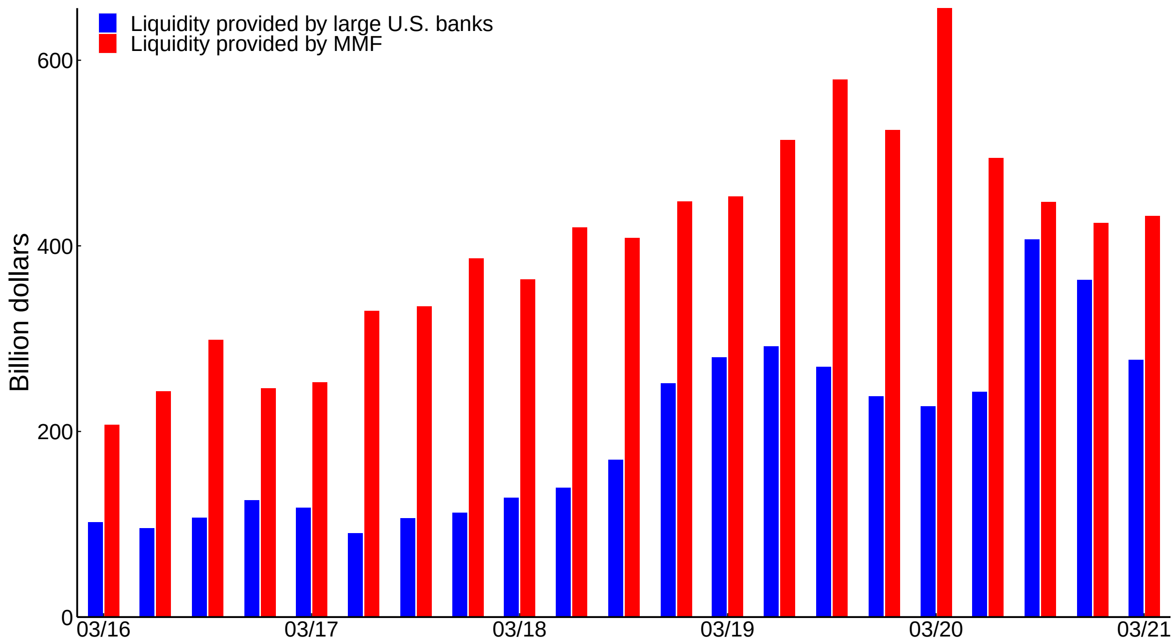

Figure 9: Liquidity provided by MMF is quarterly average of the money market mutual fund (MMF) investments in the overnight Treasury repo less MMF RRP facility usage. Liquidity provided by large U.S. banks is the quarterly average of net lending of all repo products and all tenors (reverse repos + Fed Funds lent - repos - Fed Funds borrowed). There is no exact data on the net lending of large U.S. banks in the overnight Treasury repo, but based on information from large banks' 10-Q, a lower bound is about 49% of the total net lending of all repo products. The set of large U.S. banks includes JPMorgan Chase, Bank of America, Goldman Sachs, Morgan Stanley, Citibank, Wells Fargo, and State Street. Data: FFIEC Call Reports, OFR, 10-Q.

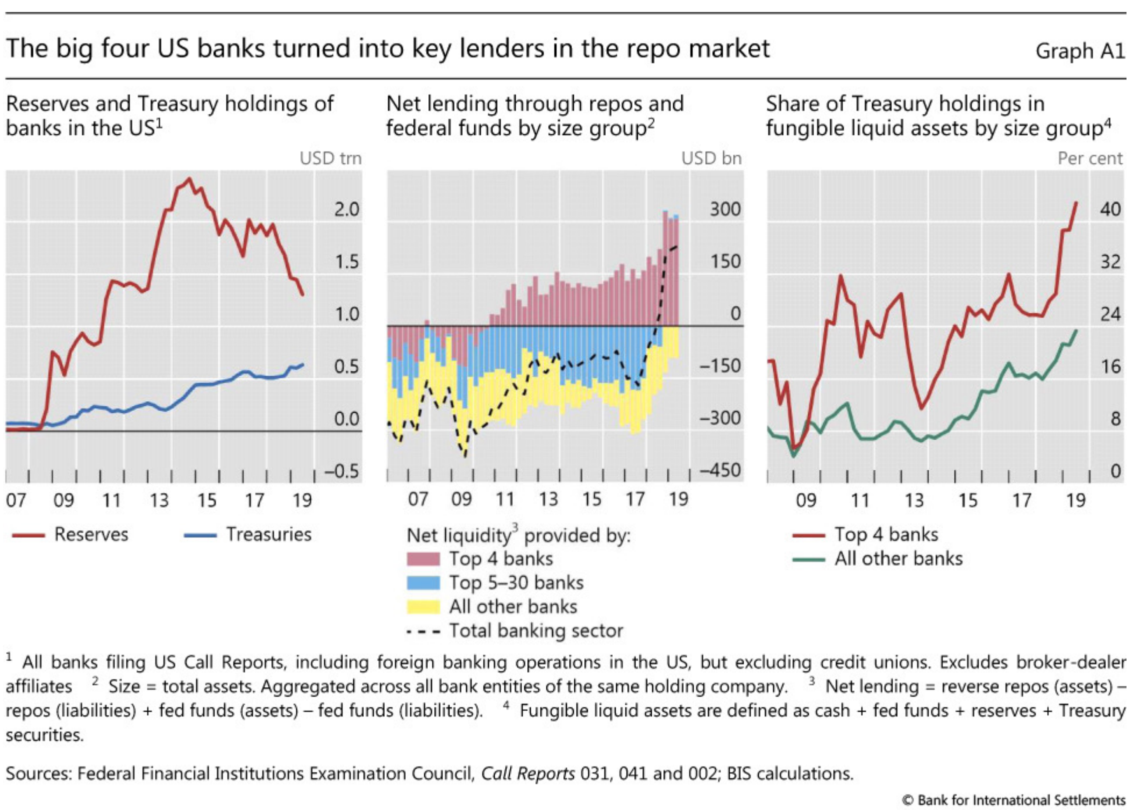

Figure 10: Source: Avalos, Ehlers and Eren (2019)

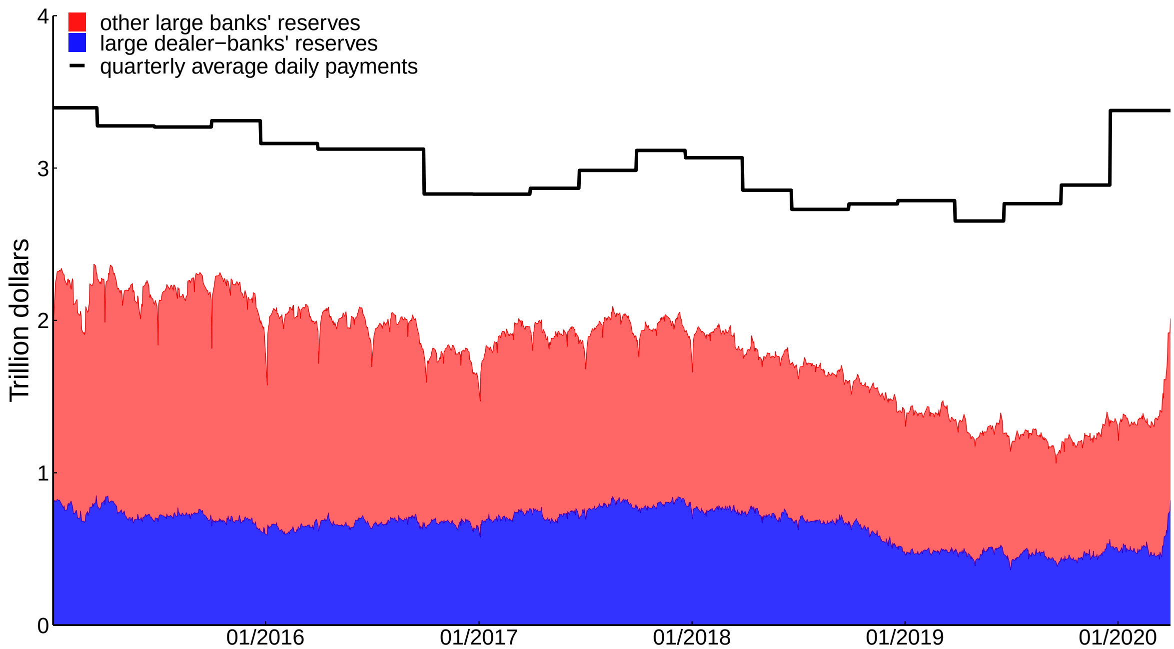

Figure 11: Quarterly average daily payments: the average daily payment value calculated every quarter by the Fedwire Funds Service. The reserve balances of the large repo-active dealer banks are shown in blue. Large repo-active dealer banks are the total reserve balances of the 10 large and repo-active account holders. Other large banks are the total reserve balances of the other large account holders of the largest 100 reserve accounts. Most of the payment activities are concentrated among the large banks (e.g. Soramäki et al. (2007)). Data: Fedwire Funds Service, FRBNY, Copeland, Duffie and Yang (2020).

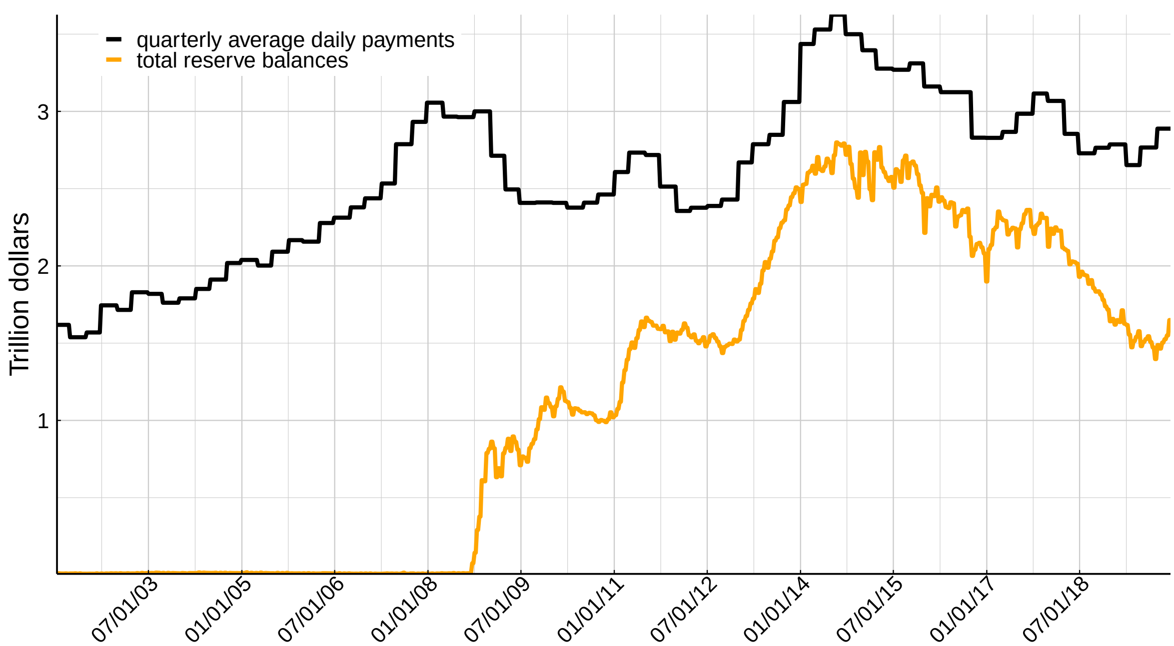

Figure 12: Quarterly average daily payments is the average daily payment value calculated every quarter by the Fedwire Funds Service. Total reserve balances average daily level of reserve balances of all depository institutions calculated every week by the Fed. Sources: Federal Reserve and Fedwire Funds Service.

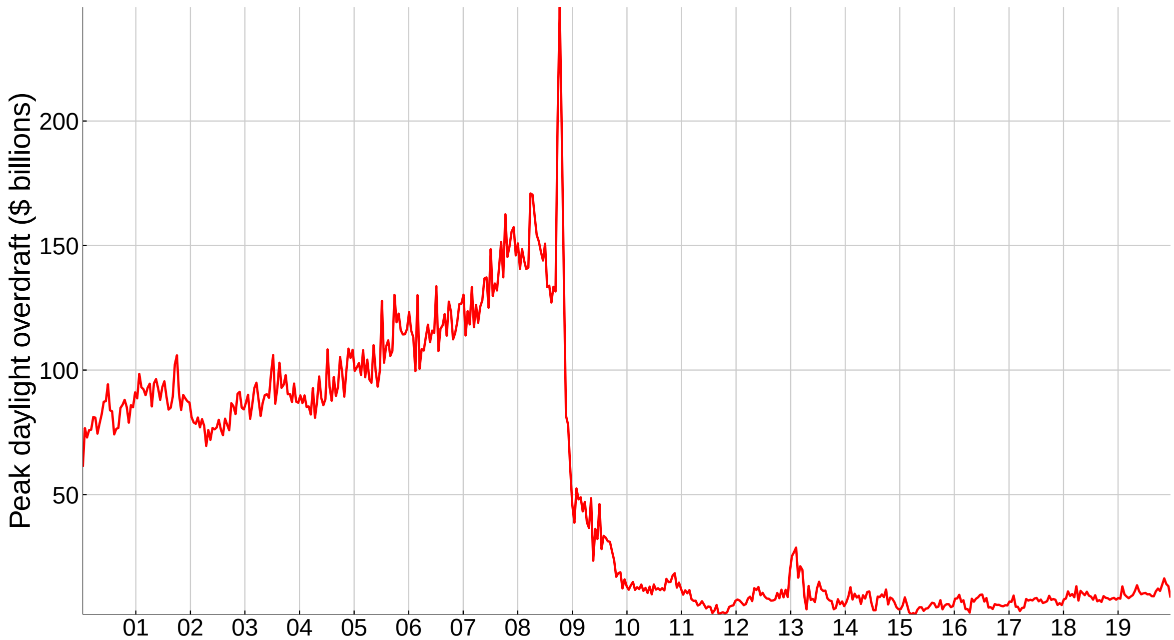

Figure 13: Peak intraday overdrafts are calculated over two-week periods and published by the Federal Reserve. The peak daylight overdraft for a given day is the greatest value reached by the sum of the daylight overdrafts for all institutions at the end of each operating minute of the day. Sources: Federal Reserve and Fedwire Funds Service.

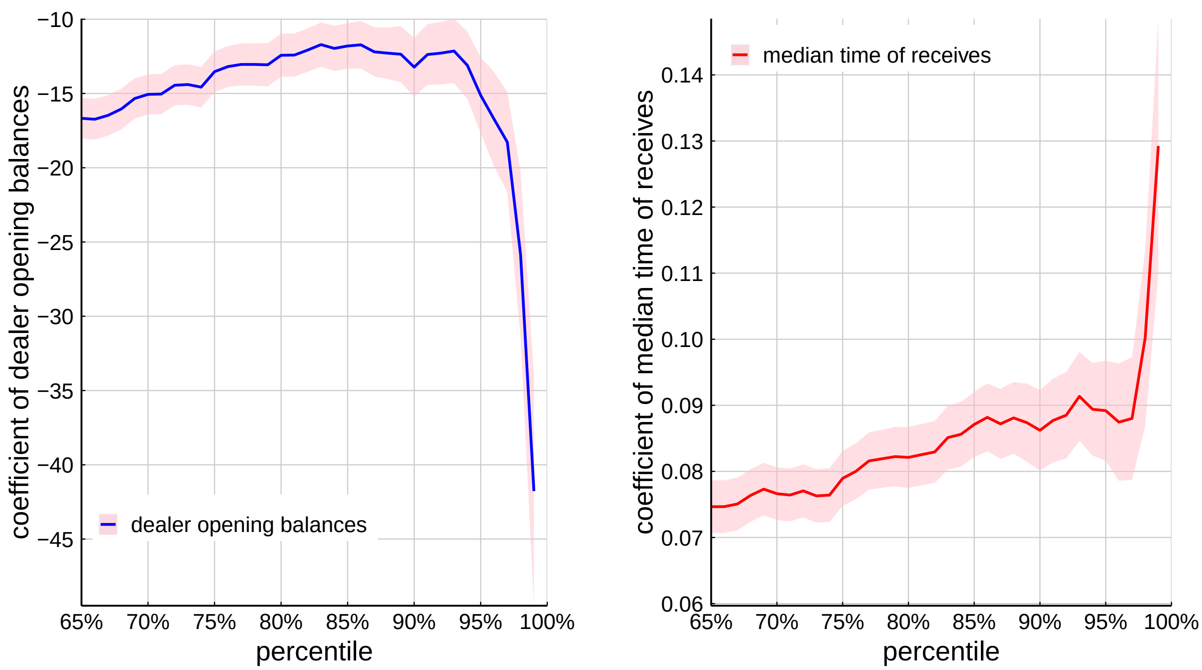

Figure 14: Quantile regression results for dealer-bank balances and payments timing. Note: Solid lines represent the coefficients of dealer-bank opening balances (unit: trillion dollars) and the median time of receives (unit: minutes) across each percentile from 65% to 99%, as specified in column (3) of Table 3. The shaded regions depict the range of one standard deviation from the point estimates. The standard deviations of dealer-bank opening balances and median time of receives in 2019 were 0.027 trillion dollars and 25.4 minutes, respectively. Source: Copeland, Duffie and Yang (2020).

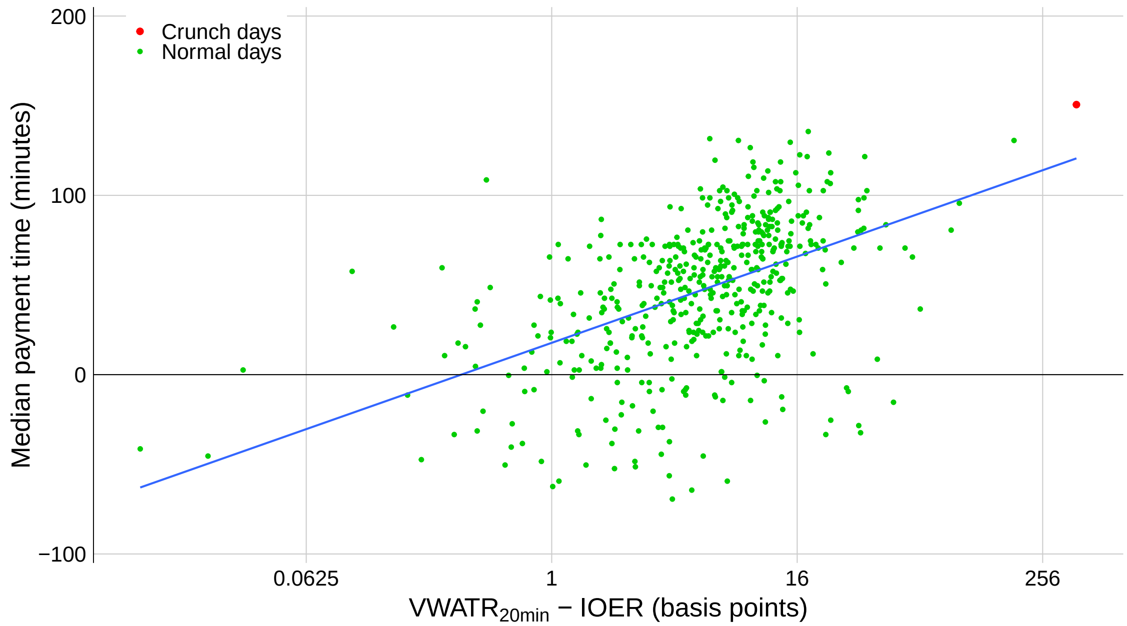

Figure 15: Payment time net of sample mean when 50% of the day's total incoming value has been received by dealer banks over Fedwire against the repo rate spread (VWATR). VWATR is the value-weighted average of the Treasury general collateral repo rate calculated from Tradition transaction data. Because of the log scale, I drop the observations for which this rate spread is negative. The upper-right dot corresponds to September 17, 2019, when repo rates had a huge spike. Clearly, payment timing had been significantly delayed on this day. Source: Copeland, Duffie and Yang (2020). Data: FRBNY and Tradition.

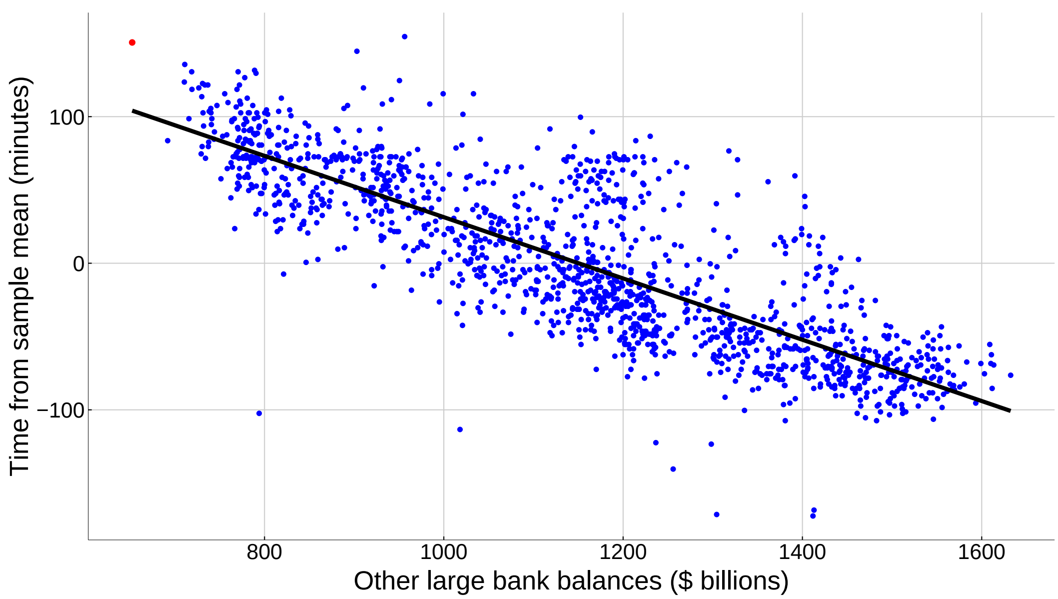

Figure 16: Non-dealer bank reserve balances and the median payment to large dealer banks. "Other large bank balances" for a given day is the total of the opening-of-day reserve balances of all accounts in our sample, except for the ten dealer banks. The payment timing measure is the half-received time of payments to the dealer banks. The date corresponding to the red dot in the upper-left corner is September 17, 2019, on which GCF− spiked to its sample-record high and the total opening balances of the other large banks reached its sample-record low. Data source: Fedwire Funds Service.

B Appendix: Institutional background

B.1 The repo market

Repo transactions are economically similar to collateralized loans. Unlike traditional collateral, repo collateral is not pledged but rather sold and then repurchased at maturity, which gives the lender greater control over the collateral. A general collateral (GC) repo is a transaction whereby the cash investor agrees to accept any security within an asset class, such as U.S. Treasuries. This paper focuses on overnight GC repo transactions collateralized by Treasuries, which constitute the largest segment of the repo market. The GC repo rates with Treasury collateral are typically free of counterparty risk and repo specialness.[40]

The repo market is among the most important global money markets. Financial institutions participating in the repo market include securities dealers, primary dealers, domestic and international banks, insurance companies, asset managers, money market funds, mutual funds, pension funds, and hedge funds. The repo market redistributes liquidity among these financial institutions, and in doing so allows other financial markets to function more efficiently. Disruptions in the repo market may undermine the efficiency and stability of the financial system.

The initial leg of an overnight repo market has "$T+0$" settlement, meaning settlement of the exchange of collateral and reserves occurs on the day the transaction is negotiated. Importantly, banks and borrowers are unable to sell other assets to provide same-day liquidity, because they are unable to obtain cash settlement for asset sales on the same day in most cases. The "$T+0$" settlement makes the repo market essential for intraday funding needs.

The repo market is critical to the implementation of monetary policy. The Federal Reserve makes heavy use of repos to manage its balance sheet and to target short-term rates, including its official target rate (i.e., the federal funds rate). After the Global Financial Crisis (GFC) of 2007-2009, the repo market supplanted the federal funds market by becoming the dominant market in which U.S. banks and dealers borrow from and lend reserves to each other. Currently, more than $5 trillion in repo products of various tenors and collateral types are traded every day.[41] One component of the market, the overnight funding market collateralized by Treasury securities and covered by SOFR, had a daily average trading volume of $1.08 trillion between January 1, 2019, and July 10, 2020. By contrast, the concurrent daily average trading volume for the federal funds market is only $0.071 trillion.[42] As a result, the Treasury-collateralized repo rate has become the most important indicator of U.S. short-term money market conditions. In addition, the Secured

[40]On occasion, one cash lender may seek a specific security as collateral in the repo market. In this case, the cash lender is willing to earn a below-market rate on the loan because the securities posted as collateral are "special," meaning they have an intrinsic value that the cash lender will attempt to monetize. This adjustment of repo rates is known as the repo specialness premium (Duffie, 1996).

[41]See Baklanova, Copeland and McCaughrin (2015) and US Repo Market Fact Sheet for more details.

[42]These estimates are based on daily-volume data from NYFed and from Fred .

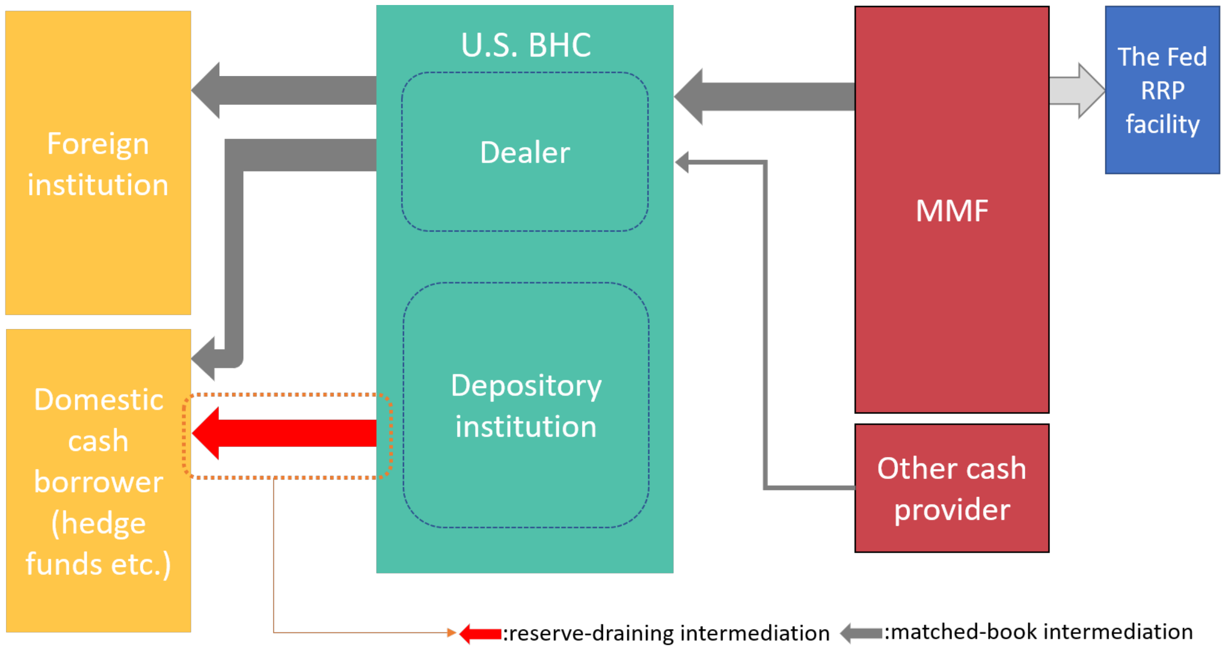

Figure 17: Stylized overview of the U.S. GC Treasury repo market. Arrows denote the flow of cash, from net cash lenders to net cash borrowers.

Overnight Financing Rate (SOFR) is replacing LIBOR as the main benchmark interest rate in U.S. money markets.[43] Therefore, an understanding of the factors that determine repo rates, especially the supply of reserves provided by the Fed, is critical to the conduct of U.S. monetary policy in the post-GFC regulatory environment.

Fig. 17 provides a stylized overview of the U.S. GC Treasury repo market. Large U.S. banks are central intermediaries in this market. On the one hand, large U.S. banks channel liquidity from ultimate cash lenders (e.g., MMFs, government-sponsored enterprises, and exchange-traded funds) to ultimate cash borrowers, including hedge funds, smaller banks, and foreign institutions.[44] In this mechanism, every dollar lent by large U.S. banks is financed by a corresponding one dollar increase in liabilities such as repo borrowing. I follow Correa, Du and Liao (2020) and call this mechanism the "matched-book intermediation" of large U.S. banks. On the other hand, large U.S. banks run down their reserve balances to provide additional liquidity in the U.S. GC Treasury repo market—so-called "reserve-draining intermediation" (Correa, Du and Liao, 2020). As demonstrated by Fig. 9, reserve-draining intermediation has played an increasingly important role in liquidity provision in short-term wholesale funding markets.

The term "repo" is usually associated with the activity of borrowing liquidity. Liquidity provision by large U.S. banks increases their reverse repo position. The market segment where large, high-quality, dealer banks borrow from U.S. money market funds is called the "triparty repo market." The market where large dealer banks lend to smaller

[43]More details of the transition from LIBOR can be found on the Alternative Reference Rates Committee's (ARRC) webpage.

[44]Foreign banks also function as dealers between lenders and borrowers, but they are primarily net cash borrowers in the GC Treasury repo market (Kahn and Olson, 2021).

dealers is called the "GCF repo market." Therefore, in a fully competitive market, the GCF- spread represents the marginal value of liquidity for large dealer banks.

The New York Fed (NY Fed) publishes SOFR as one important broad measure of the cost of borrowing cash overnight collateralized by Treasury securities. The SOFR is a volume-weighted median of transaction-level triparty repo data (collected from the Bank of New York Mellon) as well as GCF repo transaction data and data on bilateral Treasury repo transactions cleared through FICC's DVP service, which is filtered to remove a portion of transactions considered "specials." Due to its complicated composition, SOFR is not a good measure of the marginal value of liquidity for large U.S. banks, but it is highly related and typically co-moves with GCF repo rates.

B.2 The interbank payment system details

The clearing and settlement system for U.S.-dollar-denominated wholesale transactions is the largest in the world. It is highly complex and consists of a multitude of platforms that form an intricate network, connecting multiple financial institutions. The center of the network is the interbank payment system—the system that commercial banks use to send large-value or time-critical payments to each other across the accounts of the Federal Reserve, which is called the "Federal Reserve's Fedwire Funds Service" (Fedwire Funds). Fedwire Funds is a real-time gross settlement (RTGS) system and processes payments individually, immediately, unconditionally, and with finality during 22 hours of any given business day.[45] Transactions on all other platforms in the wholesale clearing and settlement system almost always involve a payment from one bank to another in the Fedwire Funds system (Bech, Martin and McAndrews, 2012). Therefore, banks face real-time demand for payment services by their clients, who wish to send money to their business counterparts who may hold accounts at other banks. Often, clients have urgent payment requests (e.g., settling foreign exchange transactions) and desire settlement by banks of potentially very large payments with minimal delay. In such a case, postponing making those payments is costly for banks because clients might either demand compensation for late settlement or take their business elsewhere in the future. In general, a bank has little control over the arrival of its customers' outgoing payment requests and the flow of its incoming funds transfers that depend on other banks' timing decisions of payment initiation. However, banks can strategically delay sending those payments (albeit delaying is costly) to smooth non-synchronized payment flows and to economize on their use of reserves throughout the day. The reason is that under post-crisis liquidity regulations and supervision, large U.S. banks appeared to have become extremely averse to allowing their intraday reserve balances to drop below a certain desired level. (See Appendix B.3 for more details.) Throughout each business day, large banks face both sizable incoming payment flows and outgoing payment requests (see Figs. 11 and 12). Therefore, they have to rely heavily on incoming payments from other banks to meet their own payment requests, and they face a serious liquidity-management problem when payment requests outbalance incoming payment flows.

[45] A more detailed description can be found here

B.3 Post-GFC liquidity regulations

I summarize some relevant liquidity rules and supervision that constrain the large dealer banks as follows:

- The Federal Reserve created the Large Institution Supervision Coordinating Committee (LISCC) supervisory program in 2010, which supervises the intraday liquidity risk of large banks. The Federal Reserve Board stated, "In 2019, LISCC liquidity supervision is focusing on the adequacy of a firm's cash-flow forecasting capabilities, practices for establishing liquidity risk limits, and measurement of intraday liquidity risk" (May, 2019 Report on Supervisory Developments).
- The Federal Reserve Board's Regulation YY, Enhanced Prudential Standards, includes rules covering intraday liquidity exposures, which state, "If the bank holding company is a global systemically important BHC, Category II bank holding company, or a Category III bank holding company, these procedures must address how the management of the bank holding company will: (i) Monitor and measure expected daily gross liquidity inflows and outflows; (ii) Manage and transfer collateral to obtain intraday credit; (iii) Identify and prioritize time-specific obligations so that the bank holding company can meet these obligations as expected and settle less critical obligations as soon as possible; (iv) Manage the issuance of credit to customers where necessary; and (v) Consider the amounts of collateral and liquidity needed to meet payment systems obligations when assessing the bank holding company's overall liquidity needs."[46]
- Resolution Liquidity Adequacy and Positioning (RLAP) under the Dodd-Frank Act-includes the intraday "resolution" liquidity requirement. The associated FDIC and Federal Reserve Board guidance states that banks must "ensure that liquidity is readily available to meet any deficits. . . Additionally, the RLAP methodology should take into account (A) the daily contractual mismatches between inflows and outflows; (B) the daily flows from movement of cash and collateral for all inter-affiliate transactions; and (C) the daily stressed liquidity flows and trapped liquidity as a result of actions taken by clients, counterparties, key FMUs,[47] and foreign supervisors, among others."

The Liquidity Coverage Ratio (LCR) is another frequently mentioned regulatory constraint that may have prevented banks from lending their excess reserves to take advantage of higher repo rates, but LCR is unlikely to present a hurdle. The LCR requires banks to hold high-quality liquid assets (HQLA) equal to a projected 30-day net cash outflow under

[46]According to the Federal Reserve Board's August 2019 Senior Financial Officer Survey, "satisfying internal liquidity stress metrics, meeting routine intraday payment flows, and meeting potential deposit outflows were important or very important determinants" of banks' holdings of excess reserves. In a related BIP survey, over three-quarters of the banks to which the Regulation YY liquidity buffer is applicable indicated this consideration to be "important" or "very important."

[47]An FMU is a designated financial market utility, such as a designated payment system or a settlement system.

stress. Excess reserves and the Treasury securities received in a reverse repo as collateral count equally as HQLA by LCR, so trading one for the other leaves a bank's HQLA unchanged. Moreover, the reverse repo is assumed to roll over 100% for 30 days, so net cash outflows are unaffected. Consequently, any bank's LCR is unchanged regardless of the amount of reserves it lends in the repo market.

Numerous industry reports and academic work have documented how the set of liquidity regulations constrained the large banks (Pozsar, 2019a,b; Younger, John and Aggarwal, 2020; Nicolae, 2020). Jamie Dimon, the Chairman and CEO of JP Morgan, commented on the September 2019 repo market disruption during J.P. Morgan's third-quarter 2019 earnings call, by saying,

. . . we have a checking account at the Fed with a certain amount of cash in it. Last year [2018] we had more cash than we needed for regulatory requirements. So when repo rates went up, we went from the checking account, which was paying IOR into repo. Obviously makes sense, you make more money. But now the cash in the account, which is still huge. It's $120 billion in the morning and goes down to $60 billion during the course of the day and back to $120 billion at the end of the day. That cash, we believe, is required under resolution and recovery and liquidity stress testing. And therefore, we could not redeploy it into repo market, which we would have been happy to do. And I think it's up to the regulators to decide they want to recalibrate the kind of liquidity they expect us to keep in that account. Again, I look at this as technical; a lot of reasons why those balances dropped to where they were. I think a lot of banks were in the same position, by the way. But I think the real issue, when you think about it, is what does that mean if we ever have bad markets? Because that's kind of hitting the red line in the Fed checking account, you're also going to hit a red line in LCR, like HQLA, which cannot redeployed either. So, to me, that will be the issue when the time comes. And it's not about JPMorgan. JPMorgan will be fine in any event. It's about how the regulators want to manage the system and who they want to intermediate when the time comes.

C Internet Appendix: Additional tables and figures

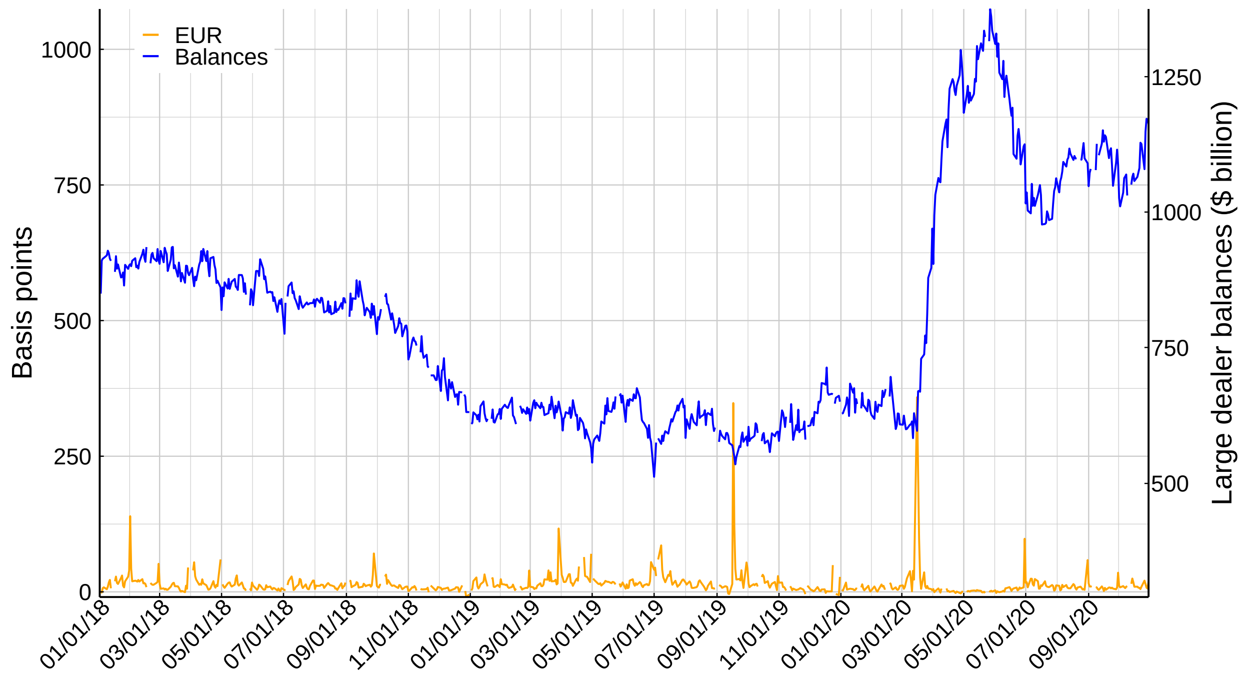

Figure 18: Reserve balances and the spread of the overnight synthetic dollar interest rates over (FX- spread). Synthetic dollar interest rate is the implied dollar interest rate in the foreign exchange (FX) swap (borrowing dollars by first borrowing in foreign currency and swapping this foreign funding for dollars, and entering into an FX forward contract to hedge the exchange-rate risk) is the interest rate paid on reserves. The reserve balances of the large repo-active banks are shown in blue (right axis). The spread of the overnight synthetic dollar funding rate by swapping the ECB deposit rate over the Fed (EUR) is shown in green (left axis). Source: Fedwire Funds Service, FRBNY, Correa, Du and Liao (2020).

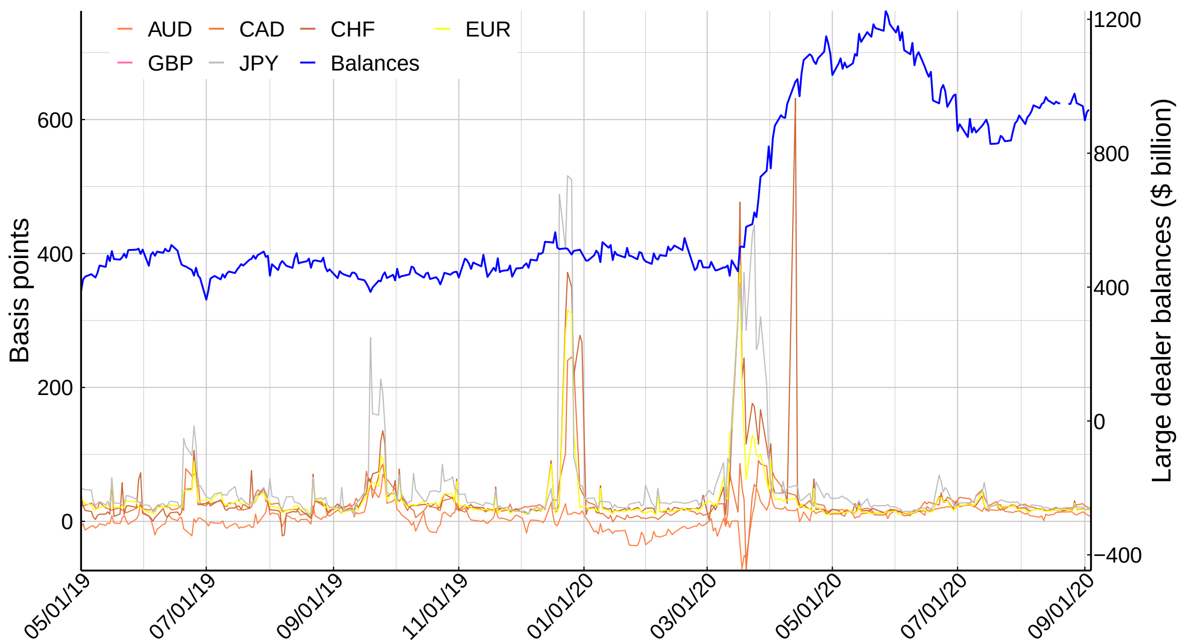

Figure 19: Reserve balances and the spread of the one-week synthetic dollar interest rates over the Overnight Index Swap (OIS) rates (left axis). For each currency, the synthetic dollar rates are calculated as the forward premium minus maturity-matched foreign currency OIS rate as in Wallen (2020). The reserve balances of the large repo-active banks are shown in blue (right axis). Source: Fedwire Funds Service, FRBNY, Bloomberg.

Note: Unlike Treasury repo loans, lending dollars in the FX markets incurs large balance sheet costs for dealer banks, especially the global systemically important banks. Due to the balance sheet constraints of international large dealer banks, the spreads of synthetic dollar interest rates over OIS (FX-OIS spread) usually spike with large magnitudes near quarter-ends and are generally more volatile (Du, Tepper and Verdelhan, 2018; Ivashina, Scharfstein and Stein, 2015; Wallen, 2020). Nevertheless, when reserve balances were low, borrowing in the FX markets became more costly.

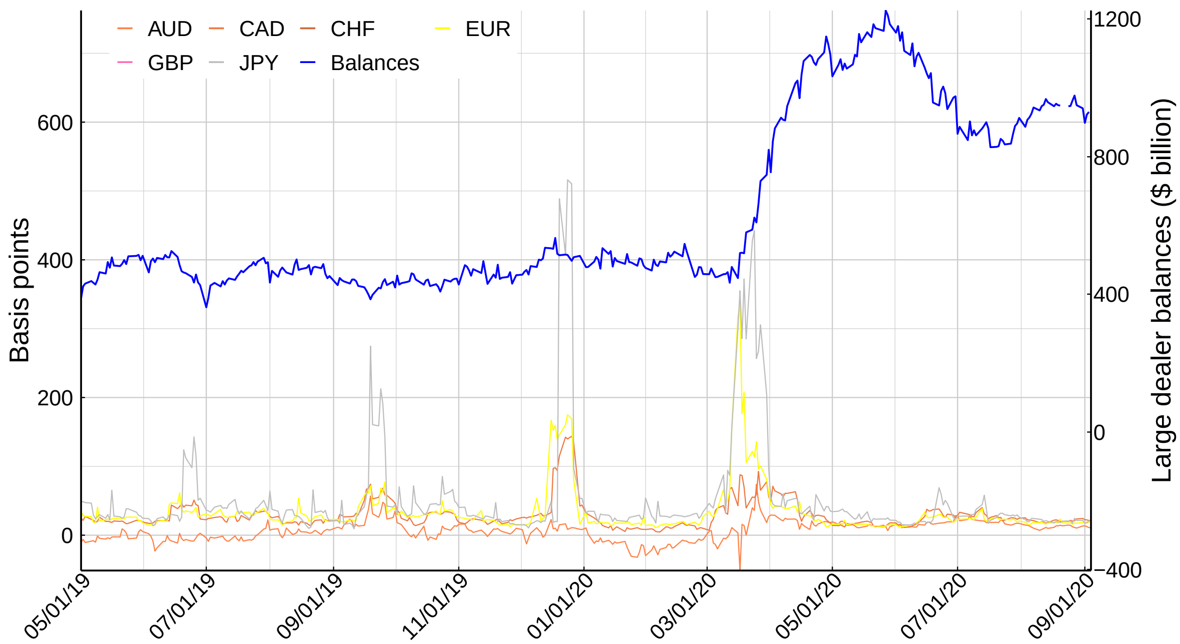

Figure 20: Reserve balances and the spread of the two-week synthetic dollar interest rates over the Overnight Index Swap (OIS) rates (left axis). For each currency, the synthetic dollar rates are calculated as the forward premium minus maturity-matched foreign currency OIS rate as in Wallen (2020).The reserve balances of the large repo-active banks are shown in blue (right axis). Source: Fedwire Funds Service, FRBNY, Bloomberg.

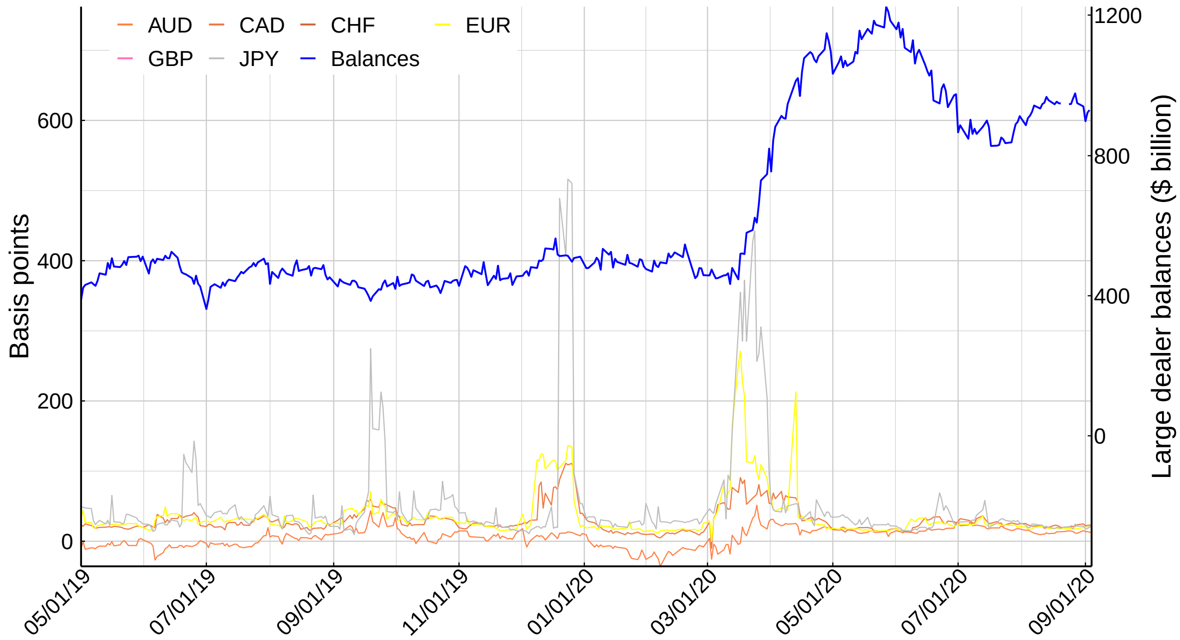

Figure 21: Reserve balances and the spread of the three-week synthetic dollar interest rates over the Overnight Index Swap (OIS) rates (left axis). For each currency, the synthetic dollar rates are calculated as the forward premium minus maturity-matched foreign currency OIS rate as in Wallen (2020). The reserve balances of the large repo-active banks are shown in blue (right axis). Source: Fedwire Funds Service, FRBNY, Bloomberg.

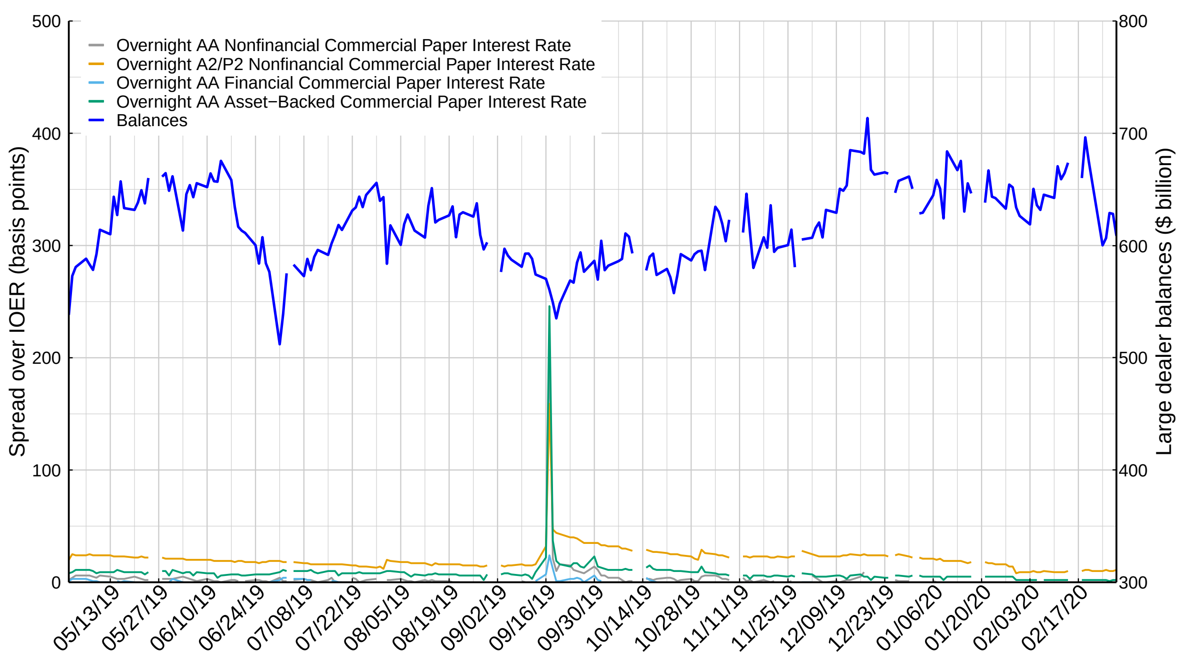

Figure 22: Reserve balances and the spreads of overnight commercial paper rates over (left axis). The reserve balances of the large repo-active banks are shown in blue (right axis). Data: Federal Reserve Board.

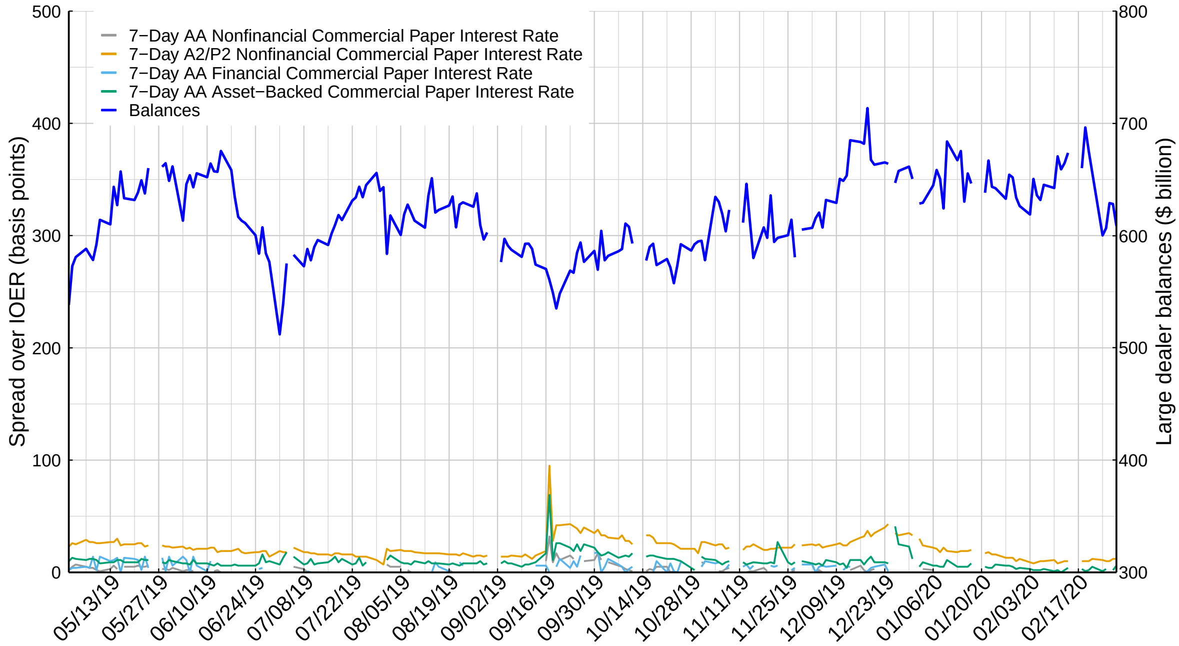

Figure 23: Reserve balances and the spreads of seven-day commercial paper rates over (left axis). The reserve balances of the large repo-active banks are shown in blue (right axis). Data: Federal Reserve Board.

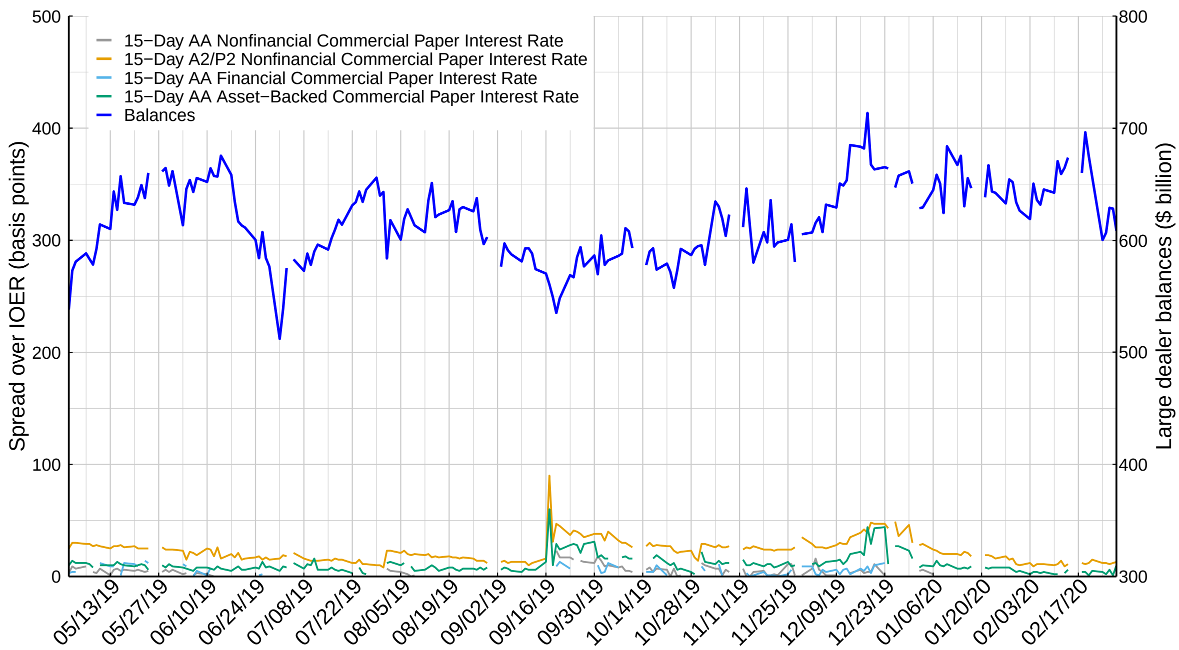

Figure 24: Reserve balances and the spreads of 15-day commercial paper rates over (left axis). The reserve balances of the large repo-active banks are shown in blue (right axis). Data: Federal Reserve Board.

D Internet Appendix: Model extensions

D.1 Monopolistic pricing

Many wholesale funding markets feature relationship trading and central-peripheral network structures. Dealer banks usually have market power over the borrowers. In this section, I extend my baseline model to encapsulate the scenario where dealer banks function as local monopolies to their respective borrowers.

As in the baseline model, in the funding market, borrower i is matched with bank i. The net cost of borrower i associated with financing amount q at funding rate r (which is endogenously determined in equilibrium), is captured by:

$$qr + \underbrace{\tfrac{\xi}{2}((D_i - q)^+)^2}_{\text{cost of reduced financing}} ,$$

For simplicity, assume the quantities D_i to be financed by borrowers have a density $f_D(x) = \lambda e^{-\lambda(x - D_{min})}$ on $[D_{min}, \infty)$.

Bank i and borrower i bilaterally negotiate the quantity-rate pair (S_i, r_i), where S_i is the quantity of reserves that bank i provides to borrower i, and r_i is the funding rate. The bilateral negotiation is modeled as a monopolistic screening model (Mussa and Rosen, 1978). Bank i acts as the local monopolist by offering a supply schedule[48] $g : \mathbb{R} \to \overline{\mathbb{R}}$, which may depend on the initial balance R_i of bank i. That is, for some measurable $G : \mathbb{R} \times \mathbb{R}_+ \to \overline{\mathbb{R}}$, bank i is willing to charge, at any quantity s chosen by the borrower, the funding rate of $g(s) = G(s, R_i)$. After observing D_i, given the supply schedule g announced by bank i, borrower i picks its desired quantity S_i by solving

$$\inf_s \quad \frac{\xi}{2}\Big((D_i - s)^+\Big)^2 + g_i(s)s, \tag{10}$$

or leaves the market without trading. To define the problem of bank i, I temporarily assume that there is a unique measurable solution $\rho(D_i, g_i)$ to (10), and that borrower i prefers obtaining $\rho(D_i, g_i)$ in funding at rate $g_i(\rho(D_i, g_i))$ to the alternative of leaving the market without trading. I show in Lemma 3 that these assumptions are satisfied in equilibrium. Having observed R_i, bank i thus chooses the supply schedule g_i to solve

$$\sup_g \quad \mathbb{E}\left[V\Big(R_i - Q - \rho(D_i, g)\Big) + g\big(\rho(D_i, g)\big)\,\rho(D_i, g) \;\Big|\; R_i\right]. \tag{11}$$

In summary, an equilibrium of the trading game consists of contingent supply schedule G and quantity of funding S_i such that given $G(\cdot, R_i)$, S_i solves the problem (10) of borrower i, and the funding schedule $G(\cdot, R_i)$ solves problem (11) of bank i. The equilibrium funding rate is $r_i = G(S_i, R_i)$.

[48] As usual, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$.

Lemma 3. *Fix any strategy* $a_i = \min((L_i + \alpha_i)^+, N_i)$ *and* $a_j = \min((L_j + \alpha_j)^+, N_j)$ *for the payment subgame. There is a unique equilibrium of the trading game which is determined by the equations*

$$
\begin{aligned}
S_i &= \mathscr{S}(D_i) \stackrel{\text{def}}{=} \inf\left\{s : \Gamma_i(R_i - Q - s, \alpha_i, \alpha_j) \geq \xi(D_i - s - \tfrac{1-F_D(D_i)}{f_D(D_i)})\right\} \\
T_i &= \mathcal{T}(D_i) \stackrel{\text{def}}{=} -\frac{\xi}{2}\Big(D_i - \mathscr{S}(D_i)\Big)^2 + \frac{\xi}{2}{D_{min}}^2 + \int_{D_{min}}^{D_i} \xi(x - \mathscr{S}(x))\,dx \\
r_i &= \frac{T_i}{S_i},
\end{aligned}
\tag{12}
$$

where F_D *is the cumulative distribution function of* D_i*.*

This formulation reflects consistent beliefs and rational expectations by bank i about the equilibrium in the subsequent payment subgame. The term $\frac{1-F_D(D_i)}{f_D(D_i)} = \lambda^{-1}$ in Eq. (12) is usually called the borrower's "information rent" in the mechanism design literature.

Lemma 4. *Fix the payment subgame equilibrium strategy profile* $\{a_i^* = \min((L_i + \alpha_i)^+, N_i)\}$*. For any* $o \in \mathbb{R}$*, and reserve balances after trading game* $L_j = R_j - Q - S_j$ *possible in equilibrium,*

$$
\begin{aligned}
L_j <\ o &\iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} <\ o \\
L_j \geq\ o &\iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} \geq\ o.
\end{aligned}
$$

Let $\Gamma_j^+(o, \alpha_j, \alpha_i) = \lim_{x \downarrow o} \Gamma_j(x, \alpha_j, \alpha_i)$*, then*

$$
\mathbb{P}(L_j \leq\ o) = \mathbb{P}(R_j - Q - D_j + \frac{\Gamma_j^+(o, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} \leq\ o) = \mathbb{E}\left[F_R\left(D_j + Q - \frac{\Gamma_j^+(o, \alpha_j, \alpha_i)}{\xi} - \lambda^{-1}\right)\right].
$$

The lemma is important because it almost bridges the endogenous state variable L_j for the payment subgame at $t = 1$ and the exogenous state variable R_j and D_j at $t = 0$ under monopolistic pricing, under which I need to revise the definition of the liquidity stress index to incorporate the information rent of borrower j:

Definition 5. The *liquidity stress index* is

$$
m = \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \leq -\frac{c}{\xi}\right) - \frac{c}{\psi} = \mathbb{E}\left[F_R\left(D_j + Q - \frac{c}{\xi} - \lambda^{-1}\right)\right] - \frac{c}{\psi},
$$

where F_R is the cumulative distribution function of R_i.
The *liquidity hoarding condition* is when

$$
m = \mathbb{E}\left[F_R\left(D_j + Q - \frac{c}{\xi} - \lambda^{-1}\right)\right] - \frac{c}{\psi} > 0. \tag{13}
$$

The *no hoarding condition* is when

$$m = \mathbb{E}\left[F_R\left(D_j + Q - \frac{c}{\xi} - \lambda^{-1}\right)\right] - \frac{c}{\psi} < 0. \tag{14}$$

For a given probability distribution of reserves satisfying a non-degeneracy condition, liquidity hoarding occurs whenever λ is sufficiently low, Q is sufficiently high, c is sufficiently low, ξ is sufficiently high, or ψ is sufficiently high.

Theorem 6. *Under the liquidity hoarding condition, there is a unique equilibrium. In this equilibrium, bank i hoards liquidity and pays $a_i^* = \min(L_i^+, N_i)$ in the payment subgame. The marginal value of liquidity functions are the same for both banks $\Gamma_i(y,0,0) = \Gamma_j(y,0,0) = \Gamma(y,0,0)$ such that*

1. *When $y > 0$, $\Gamma(y,0,0) = c(1 - F_N(y))$;*
2. *When $y \le 0$, $\Gamma(y,0,0) = \psi\left(F_N(-y) + (1 - F_N(-y))F_{RD}(-\lambda^{-1} - y - \frac{\Gamma(-y,0,0)}{\xi})\right)$.*

Theorem 7. *Under no liquidity hoarding, there always exists at least one equilibrium. Any equilibrium must be symmetric (in the sense that $\alpha_i = \alpha_j = \alpha$) with pure payment strategy $a_i^* = \min(N_i, (L_i + \alpha)^+)$ for some $\alpha > N_{min}$. The marginal value of liquidity functions are the same for both banks $\Gamma_i(y,\alpha,\alpha) = \Gamma_j(y,\alpha,\alpha) = \Gamma(y,\alpha,\alpha)$. In addition, α and Γ solve a system of integral equations:*

$$\mathbb{P}\left(R_i - D_i - Q \le -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right) + \mathbb{P}(N_i \le \alpha)\left(1 - \mathbb{P}\left(R_i - D_i - Q \le -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)\right) = \frac{c}{\psi};$$

$$\Gamma(y,\alpha,\alpha) = \begin{cases} \psi \int_{n \in (y^+, (y+\alpha))} F_N(n-y) + (1 - F_N(n-y))F_{RD}(-\lambda^{-1} + n - y - \alpha - \frac{\Gamma(n-y-\alpha,\alpha,\alpha)}{\xi}) \; dF_N(n) & \\ \qquad + \int_{n \in [(y+\alpha),\infty)} c \, dF_N(n), & \forall y > -\alpha; \\ \psi\left(F_N(-y) + (1 - F_N(-y))F_{RD}(-\lambda^{-1} - y - \alpha - \frac{\Gamma(-y-\alpha,\alpha,\alpha)}{\xi})\right) & \forall y \le -\alpha. \end{cases}$$

I find the equilibrium is unique under the same technical conditions as in Section 4.2.

Theorem 8. *Assume $N_i - N_{min}$ $(i = 1,2)$ is exponentially distributed with parameter λ_N, and F_{RD} is differentiable with density function f_{RD}. Let $f_{RD}^m = \sup\{f_{RD}(t) : t \le 0\}$. If $\frac{\sqrt{2}e\xi}{\psi} > f_{RD}^m$, the equilibrium is unique under the no hoarding condition.*

Theorem 6 and Theorem 8 imply the equilibrium is unique under both the liquidity hoarding condition and the no hoarding condition. Once the payment subgame and the marginal value for liquidity are determined, the equilibrium funding rate can be solved directly:

Theorem 9. *Given realizations of R_i and D_i, equilibrium trading quantity $S_i = \mathscr{S}(D_i)$ is* *determined by Eq. (12) in Lemma 3. The equilibrium funding rate r_i is*

$$r_i = \frac{\int_{R_i-Q-\mathscr{S}(D_i)}^{R_i-Q-\mathscr{S}(D_{min})} \Gamma(y,\alpha,\alpha)\,dy + \mathcal{T}(D_{min}) - \xi\lambda^{-1}\mathscr{S}(D_{min})}{\mathscr{S}(D_i)} + \xi\lambda^{-1},$$

where the marginal value $\Gamma(y,\alpha,\alpha)$ of reserves is solved explicitly in Theorem 6 and Theorem 7, case by case.

Theorem 7 and Theorem 9 state that when banks expect other banks to have abundant opening reserves, they have a better incentive to send more payments early in the day and to lend more liberally in the funding market. Theorem 6 and Theorem 9 predict that when market conditions (including the probability distribution of reserve levels) change slightly, yet enough to trigger the hoarding condition Eq. (13), the marginal value of liquidity Γ can jump up from $\Gamma(R_i - Q - S_i, \alpha, \alpha)$ to $\Gamma(R_i - Q - S_i, 0, 0)$, causing short-term funding rates to spike:

Theorem 10. *Fix some outcome ζ of beginning reserve balances R_i and a quantity $\mathcal{S}^*$ traded in the funding market. The equilibrium funding rate r^* jumps up as a function of the liquidity stress m at the threshold $m = 0$ that triggers liquidity hoarding. More specifically, there exists some $\delta(\zeta, \mathcal{S}^*) > 0$ such that*

$$\lim_{\epsilon^m \downarrow 0} r^*(\zeta, \mathcal{S}^*, \epsilon^m) - r^*(\zeta, \mathcal{S}^*, -\epsilon^m) > \delta(\zeta, \mathcal{S}^*),$$

provided that the sets of macroeconomic conditions $\mathcal{M}_C{}^{\epsilon^m}$ are mutually close to each other with respect to the liquidity stress index.

D.2 General case: n dealer banks

In this section, I present and solve the model for $n > 2$ dealer banks. For simplicity, assume n dealer banks are symmetric.[49] Assume the random variables of total payment needs for each bank N_1, N_2, $\cdots$ N_n are identically and independently distributed according to a probability distribution function (pdf) $f_N(\cdot)$ on the support $[N_{min}, \infty)$ for some $N_{min} > 0$. In addition, assume the initial reserve balances for each bank $R_1, R_2, \cdots, R_n$ are identically and independently distributed with pdf $f_R(\cdot)$. Similarly, assume the borrowing demands from each short-term borrower D_1, D_2, $\cdots$ and D_n are identically and independently distributed with pdf $f_D(\cdot)$. Let $F_{RD}(x) = \mathbb{P}(R_i - D_i - Q \leq x)$. Under stated conditions, F_{RD} is differentiable with pdf f_{RD}.

Suppose bank i makes an early payment $a_{i,j}$ to bank j, and gets incoming early payment $a_{j,i}$ from bank j. Then in the payment subgame the cost to bank i associated with

[49] Although numerically solving the model with asymmetric banks is possible, the lack of granular data prevents us from examining empirically the implications of asymmetric banks.

payment timing is

$$\psi(L_i - \sum_{j\neq i} a_{i,j} + \sum_{j\neq i} a_{j,i})^- + c(N_i - \sum_{j\neq i} a_{i,j})^+ = \psi(L_i - a_i + a_{-i})^- + c(N_i - a_i)^+,$$

where

$$a_i \overset{\text{def}}{=} \sum_{j\neq i} a_{i,j}, \qquad \text{and} \qquad a_{-i} \overset{\text{def}}{=} \sum_{j\neq i} a_{j,i},$$

and L_i is the reserve balances of bank i after the lending in the funding market. Clearly, bank i is indifferent regarding how to split its total outgoing payment to other banks as long as the total payment a_i is the same. To simplify the analysis, I assume each of the large dealer banks has the same business relationship with all other dealer banks. (Soramäki, Bech, Arnold, Glass and Beyeler (2007) show 25 large banks form a densely connected sub-graph, or clique, in the payment network of the Fedwire system.) Therefore, $a_{i,j} = a_{i,k} = \frac{a_{-i}}{n-1}$ for any $j, k \neq i$.

Given the payment strategy a_{-i} of all other banks, bank i chooses a_i to optimize the conditional expected payoff in the payment subgame

$$U(L_i, N_i) = \mathbb{E}[-\psi\,(L_i - a_i + a_{-i})^- - c\,(N_i - a_i)^+ \;\big|\; N_i, L_i]. \tag{15}$$

Lemma 5. *Suppose that either $\prod_{j\neq i} \mathbb{P}(L_j \leq 0) > \frac{c}{\psi}$ for all i, or $\prod_{j\neq i} \mathbb{P}(L_j \leq 0) < \frac{c}{\psi}$ for all i. Then there is a unique Perfect Bayesian payment game equilibrium. In this equilibrium, each bank i chooses the payment $a_i^* = \min((L_i + \alpha_i)^+, N_i)$, where (α_i) solves*

$$\alpha_i = \inf\left\{\vartheta \geq 0 : \ \mathbb{P}(\sum_{j\neq i} \min((L_j + \alpha_j)^+, N_j) \leq \vartheta) \geq \frac{c}{\psi}\right\} \tag{16}$$

In particular, if L_i are i.i.d. distributed, then there is a Perfect Bayesian payment game equilibrium of the form $a_i^* = \min((L_i + \alpha)^+, N_i)$, for some constants α such that

$$\alpha = \inf\left\{ \vartheta \geq 0 : \quad \mathbb{P}(\sum_{j\neq i} \min((L_j + \vartheta)^+ - \vartheta, N_j - \vartheta) \leq 0) \geq \frac{c}{\psi} \right\}.$$

The equilibrium is unique, except for the knife-edge case, where $\mathbb{P}(L_j \leq 0) = \sqrt[n-1]{c/\psi}$.

The funding market is modeled as an OTC market where bank i and borrower i are matched and are price-takers, as in Section 4.2: The cost to borrower i of borrowing S_i at funding rate r_i is $S_i r_i + (D_i - S_i)^+ r + \frac{\xi}{2}((D_i - S_i)^+)^2$. Bank i's payoff for lending S_i at funding rate r_i is $S_i(r_i - r) + V(R_i - S_i)$. Because I assume the distribution functions of R_i and D_i are atomless, the equilibrium funding rate r^* and trading quantity S_i^* satisfy

$$V'(R_i - S_i^*) = r_i^* - r = \xi(D_i - S_i^*),$$

where the marginal value of liquidity $\Gamma = V'$ depends on the strategy profile of all other banks in the payment subgame and can be calculated directly:

Lemma 6. *Given the payment subgame strategy profile* $\{a_i = \min((L_i + \alpha_i)^+, N_i)\}$ *and some joint probability distribution for* $\{L_i\}$, *let* $\alpha_{-i} = (\alpha_j)_{j \neq i}$. *Then, the marginal value of liquidity function for bank* i *to be the function* $\Gamma_i : \mathbb{R} \times \mathbb{R}^{+n} \to \mathbb{R}^+$ *such that*

$$\Gamma_i(y, \alpha_i, \alpha_{-i}) = \begin{cases} \int\limits_{\eta \in (y^+, (y+\alpha_i))} \psi \mathbb{P}(\frac{\sum_{j \neq i} \min((L_j+\alpha_j)^+, N_i)}{n-1} \leq \eta - y) \, dF_N(\eta) + \int\limits_{\eta \in [(y+\alpha_i), \infty)} c dF_N(\eta), & \forall y > -\alpha_i \\ \psi \mathbb{P}(\frac{\sum_{j \neq i} \min((L_j+\alpha_j)^+, N_j)}{n-1} \leq -y), & \forall y \leq -\alpha_i. \end{cases}$$

With identical proof as in the two-banks case, we have

Lemma 7. *Fix the payment subgame equilibrium strategy profile* $\{a_i^* = \min((L_i + \alpha_i)^+, N_i)\}$. *For any* $o \in \mathbb{R}$, *and reserve balances after trade* $L_j = R_j - Q - S_j$ *possible in equilibrium,*

$$L_j < \; o \iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_{-j})}{\xi} < \; o$$
$$L_j \geq \; o \iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_{-j})}{\xi} \geq \; o.$$

Let $\Gamma_j^+(o, \alpha_j, \alpha_{-j}) = \lim_{x \downarrow o} \Gamma_j^+(x, \alpha_j, \alpha_{-j})$, *then* $\mathbb{P}(L_j \leq \; o) = \mathbb{P}(R_j - Q - D_j + \frac{\Gamma_j^+(o, \alpha_j, \alpha_{-j})}{\xi} \leq \; o)$.

Recall that, by assumption, $\{R_i\}$ and $\{D_i\}$ are i.i.d. As in the two-banks case, Lemma 7 motivates the following definitions:

Definition 6. The *liquidity hoarding condition for* n *banks* is

$$\mathbb{P}(R_i - D_i - Q \leq -\frac{c}{\xi}) > (\frac{c}{\psi})^{1/(n-1)}. \tag{17}$$

The *no hoarding condition for* n *banks* is

$$\mathbb{P}(R_i - D_i - Q \leq -\frac{c}{\xi}) < (\frac{c}{\psi})^{1/(n-1)}. \tag{18}$$

Note the *liquidity hoarding condition for* n *banks* is less stringent than the *liquidity hoarding condition* for two banks, due to the diversification benefits. Each bank receives incoming payments from $n - 1 > 1$ other banks, so it is not as concerned about the liquidity condition of any particular bank. When $n \to \infty$, bank i knows it will receive a positive amount of incoming payments: $a_{-i} = \frac{\sum_{j \neq i} \min((L_j+\alpha_j)^+, N_j)}{n-1} > 0$ almost surely if $\mathbb{P}(R_j - D_j - Q \leq -\frac{c}{\xi}) < 1$. In this case, bank i will never hoard liquidity in the payment game. In other words, *liquidity hoarding condition for* n *banks* will never hold as $n \to \infty$ when $\mathbb{P}(R_j - D_j - Q \leq -\frac{c}{\xi}) < 1$.

Apparently, it is useful to consider the probability distribution of a_{-i} for bank i. Fix some payment subgame equilibrium strategy profile $\{a_i^* = \min((L_i + \alpha_i)^+, N_i)\}$. By assumption, the joint probability distribution of $\{a_i^*\}$ is common knowledge for all banks. Let

$$F_{a_j}(x) \overset{def}{=} \mathbb{P}(\min((L_j+\alpha_j)^+, N_j) \leq x).$$

Lemma 7 implies that for any $x \geq 0$,

$$F_{a_j}(x) = F_N(x) + (1-F_N(x))\mathbb{P}(R_j - D_j - Q \leq x - \alpha_j - \frac{\Gamma_j(x-\alpha_j, \alpha_j, \alpha_{-j})}{\xi}).$$

Under the stated conditions, when $x > 0$, $F_{a_j}(x)$ is differentiable. Let $f_{a_j}(\cdot)$ be the pdf for a_j. Direct calculation gives

Lemma 8. *Given some payment subgame strategy profile $\{a_i = \min((L_i+\alpha_i)^+, N_i)\}$ and some joint probability distribution for $\{L_i\}$, then for any $x \geq N_{min}$,*

$$\begin{aligned} f_{a_j}(x) = \frac{dF_{a_j}(x)}{dx} =& (1-F_N(x))f_{RD}(x-\alpha_j - \frac{\Gamma_j(x-\alpha_j)}{\xi})(1-\frac{\Gamma_j'(x-\alpha_j)}{\xi}) \\ &+ f_N(x)(1-F_{RD}(x-\alpha_j-\frac{\Gamma_j(x-\alpha_j)}{\xi})), \end{aligned}$$

and for any $0 < x < N_{min}$,

$$f_{a_j}(x) = f_{RD}(x-\alpha_j - \frac{\Gamma_j(x-\alpha_j)}{\xi})(1-\frac{\Gamma_j'(x-\alpha_j)}{\xi}).$$

For any $x \geq 0$,

$$F_{a_{-i}}(x) \overset{def}{=} \mathbb{P}(\frac{\sum_{j\neq i}\min((L_j+\alpha_j)^+, N_j)}{n-1} \leq x) = \prod_{j\neq i} F_{a_j}(0) + \int_0^{(n-1)x} f_{\sum_{j\neq i} a_j}(t)dt,$$

where $f_{\sum_{j\neq i} a_j}(\cdot)$ is the pdf for the random variable $\sum_{j\neq i} a_j$ that can be calculated using the convolution operation on $\{f_{a_j}\}_{j\neq i}$.

With Lemma 8, the equilibrium can be characterized by the following theorems:

Theorem 11. *Under the liquidity hoarding condition for n banks, there is a unique equilibrium, in which bank i pays $a_i^* = \min(N_i, L_i^+)$ in the payment subgame. The marginal value of liquidity functions are the same for all banks $\Gamma_i(y,0,0) = \Gamma(y,0,0)$. In addition,*

1. *When $y > 0$, $\Gamma(y,0,0) = c(1-F_N(y))$;*
2. *When $y \leq 0$, $\Gamma(y,0,0) = \psi F_{a_{-i}}(-y)$.*

Theorem 12. *Under no liquidity hoarding, there always exists at least one equilibrium. Any equilibrium must be symmetric with pure payment strategy $a_i^* = \min(N_i, (L_i+\alpha)^+)$ for some*

$\alpha > N_{min}$. *The marginal value of liquidity functions are the same for both banks* $\Gamma_i(y,\alpha,\alpha) = \Gamma_j(y,\alpha,\alpha) = \Gamma(y,\alpha,\alpha)$. *In addition,* α *and* Γ *solve the following system of integral equations:*

$$F_{a_{-i}}(\alpha) = \frac{c}{\psi}$$
$$\Gamma(y,\alpha,\alpha) = \begin{cases} \int_{n\in(y^+,(y+\alpha))} \psi F_{a_{-i}}(n-y)\, dF_N(n) + \int_{n\in[(y+\alpha),\infty)} c\, dF_N(n), & \forall y > -\alpha \\ \psi F_{a_{-i}}(-y), & \forall y \leq -\alpha. \end{cases} \tag{19}$$

Finally, Theorems 11 and 12 imply the Theorems 4 and 5 hold for the $n-$banks case. The proofs are almost identical to the two-banks case.

D.3 A quantitative model without strategic complementarity

To highlight the important role of strategic complementarity, I examine an analogous model without strategic complementarity. I follow the same data processing approach detailed in Section 8, presuming a linear impact of Treasury issuances on large dealer banks' reserve balances and posits that large non-dealer banks' early payments are linearly related to their total reserve balances. From the observed dealer balance R_D, I calculate:

$$R_i = \tfrac{1}{10}(R_D - \underbrace{E_I^{NC} T_I}_{\substack{\text{Impact from}\\ \text{Treasury issuance}}} + \underbrace{E_O^{NC} R_O)}_{\substack{\text{Early incoming payment}\\ \text{from other banks}}} - Q^{NC},$$

where E_I^{NC}, E_O^{NC} and Q^{NC} are constants to be estimated, representing the payment effects of Treasury issuance, early incoming payments from non-dealer banks, and a normalization constant, respectively.

Following the benchmark model described in Section 4.2, I maintain the timeline and the assumption of competitive funding markets. After deploying S_i in the funding market, bank i's remaining reserve balance for the payment subgame is $R_i - S_i$. For the exercise in this section, I exclude the element of strategic complementarity by assuming that bank i does not strategically factor in bank j's optimal payment strategy as outlined in Section 4.1. Instead, bank i presumes bank j's early payment to be a constant fraction of bank j's projected total reserve balances, less expected loans in the funding markets. Specifically, bank i non-strategically posits $a_j^* = [\min(\mathbb{E}^i[R_j - S_j] + E_R^{NC}, N_j)]^+$ and $\mathbb{E}^i[S_j] = E_D^{NC} T_D$, where T_D indicates the total outstanding Treasuries (also normalized by dividing by 10), and E_R^{NC}, E_D^{NC} are constants determined through estimation. Consistent with Section 4.2, I posit reserves R_D are symmetrically distributed among large dealer banks, leading bank i's simple belief that $\mathbb{E}^i[R_j] = R_i$.

For the payment subgame, bank i can make any non-negative payment $a_i \leq N_i$ at time 1, deferring any remaining payment to time 2. Given bank i's perception of bank j's payment strategy a_j^*, bank i seeks to optimize its conditional expected payoff by selecting

a_i:

$$U = \mathbb{E}^i[-\psi^{NC}(R_i - S_i - a_i + a_j^*)^- - c^{NC}(N_i - a_i)^+ \mid N_i, R_i, S_i]$$
$$= -\psi^{NC}(R_i - S_i - a_i + [\min(R_i - E_D^{NC}T_D + E_R^{NC}, N_j)]^+)^- - c^{NC}(N_i - a_i)^+.$$

Following the approach outlined in Section 8, I model the deviation of bank i's payment obligation from its minimum, $N_i - N_{\min}$, as following an exponential distribution with parameter λ_N, where both λ_N and $N_{\min}$ are estimated from the payment data. N_j and N_i have the same distribution. I keep the assumption that $\psi^{NC} > c^{NC}$, so the optimal strategy for bank i is thus $a_i^* = \min((R_i - S_i + \alpha^*)^+, N_i)$ for some α^* such that

$$\alpha^* = \inf\left\{ \vartheta \geq 0 : \mathbb{P}([\min(R_i - E_D^{NC}T_D + E_R^{NC}, N_j)]^+ \leq \vartheta) \geq \frac{c^{NC}}{\psi^{NC}} \right\}.$$

Let $N^C = N_{min} + \frac{\log(\psi^{NC}) - \log(\psi^{NC} - c^{NC})}{\lambda_N}$, so $\mathbb{P}(N_j \leq N^C) = \frac{c^{NC}}{\psi^{NC}}$. Consequently, α^* is characterized as follows:

$$\alpha^* = \min((R_i - E_D^{NC}T_D + E_R^{NC})^+, N^C).$$

The continuation value of bank i for reserve balances at the beginning of the payment game before observing its payment obligation N_i is given by:

$$V(R_i - S_i) = \mathbb{E}^i[-\psi^{NC}(R_i - S_i - a_i^* + a_j^*)^- - c^{NC}(N_i - a_i^*)^+ \mid R_i, S_i, T_D]$$
$$= \mathbb{E}[-\psi^{NC}\left(a_i^* - (R_i - S_i) - \min(\mathbb{E}^i[R_j - S_j] + E_R^{NC}, N_j)\right)^+ - c^{NC}(N_i - a_i^*)^+ \mid R_i, S_i, T_D].$$

A direct calculation of the left-hand derivative of V with respect to reserve balances, denoted $\Gamma(y) \equiv V'_-(y)$, yields

$$\Gamma(y) = \lim_{x \uparrow y} \frac{V(x) - V(y)}{x - y}$$
$$= \begin{cases} \int_{n \in [y^+, (y+\alpha^*))} \psi^{NC}\mathbb{P}(a_j^* \leq n - y) dF_N(n) + \int_{n \in [(y+\alpha^*), \infty)} c^{NC} dF_N(n) & \text{when} \quad y + \alpha^* > 0; \\ \psi^{NC}\mathbb{P}(a_j \leq -y) & \text{otherwise.} \end{cases}$$

$\Gamma(y)$ is the marginal value of liquidity, which determines the whole funding rates in a competitive market. Throughout 2019, large banks consistently complied with liquidity regulations every day, ensuring $R_i - S_i - a_i > 0$ within my dataset (Afonso, Cipriani, Copeland, Kovner, La Spada and Martin, 2020b). Banks also understand that their counterparts are also very unlikely to breach liquidity requirements, implying $\mathbb{E}^i[R_j - S_j] > 0$. My analysis, therefore, concentrates on parameters ensuring $R_i - E_D^{NC}T_D + E_R^{NC} \geq 0$ for the training sample. We further simplify the marginal value $\Gamma(y)$ when $y \geq 0$, because this is the relevant case as empirically $y = R_i - S_i \geq 0$ throughout my sample: Given the empirical observation that $y = R_i - S_i \geq 0$ consistently across the dataset, our analysis will concentrate on scenarios where $y \geq 0$. This allows for a further simplification of the marginal value $\Gamma(y)$, detailed as follows:

$$\Gamma(y) = \int\limits_{n\in[y,(y+\alpha^*))} \psi^{NC}\mathbb{P}(\min(R_i - E_D^{NC}T_D + E_R^{NC}, N_j) \leq n - y)dF_N(n) + \int\limits_{n\in[(y+\alpha^*),\infty)} c^{NC}dF_N(n)$$

Since $n-y < \alpha^* = \min(R_i - E_D^{NC}T_D + E_R^{NC}, N^C) \leq R_i - E_D^{NC}T_D + E_R^{NC}$ when $n \in [y,(y+\alpha^*))$,

$$\begin{aligned}\Gamma(y) =& \int\limits_{n\in[y,(y+\alpha^*))} \psi^{NC}\mathbb{P}(N_j \leq n - y)dF_N(n) + \int\limits_{n\in[(y+\alpha^*),\infty)} c^{NC}dF_N(n)\\ =&\psi^{NC}\lambda_N \int\limits_{y+\min(N_{min},\alpha^*)}^{y+\alpha^*} (1-e^{-\lambda_N(n-y-N_{min})})e^{-\lambda_N(n-N_{min})}dn + c^{NC}e^{-\lambda_N(y+\alpha^*-N_{min})^+}\\ =&\frac{1}{2}\psi^{NC}e^{-\lambda_N(y)}(e^{-\lambda_N(\alpha^*-N_{min})^+}-1)^2 + c^{NC}e^{-\lambda_N(\alpha^*-N_{min})^+}\end{aligned} \tag{20}$$

Following Section 8, this model is estimated within the context of the GCF Treasury repo market, using a training sample from January 3, 2019, to August 31, 2019. In competitive funding markets, the repo rates are determined by the marginal continuation value for large dealer banks in the payment subgame. Thus, according to Eq. (20), the empirical relationship under this model—excluding strategic complementarity—for day t can be expressed as follows:

$$\begin{aligned}&(GCF-)^t\\ =&\frac{1}{2}\psi^{NC}e^{-\lambda_N(R_i-S_i)}(e^{-\lambda_N(\alpha^*-N_{min})^+}-1)^2 + c^{NC}e^{-\lambda_N(\alpha^*-N_{min})^+} + \overline{\varrho^{NC}}E^t_{month} + \overline{\vartheta^{NC}} + \epsilon^t_r\\ =&\frac{1}{2}\psi^{NC}e^{-\lambda_N(\frac{R_D-E_I^{NC}T_I+E_O^{NC}R_O}{10}-Q^{NC}-S_i)}(e^{-\lambda_N(\min((\frac{R_D-E_I^{NC}T_I+E_O^{NC}R_O}{10}-Q^{NC}-E_D^{NC}T_D+E_R^{NC})^+,N^C)-N_{min})^+}-1)^2\\ &+c^{NC}e^{-\lambda_N(\min((\frac{R_D-E_I^{NC}T_I+E_O^{NC}R_O}{10}-Q^{NC}-E_D^{NC}T_D+E_R^{NC})^+,N^C)-N_{min})^+} + \overline{\varrho^{NC}}E^t_{month} + \overline{\vartheta^{NC}} + \epsilon^t_r.\end{aligned}$$

Here, E^t_{month} represents the month-end indicator, while $\overline{\varrho^{NC}}$ and $\overline{\vartheta^{NC}}$ are constants to be estimated. The parameter Q^{NC}'s impact is effectively encapsulated by adjustments in ψ^{NC} and E_R^{NC}, rendering it unidentifiable on its own. Therefore, in alignment with Section 8, I pick the value of $Q^{NC} = Q = \$16.19$ billion dollars. Importantly, given that the estimation procedures can freely adjust ψ^{NC} and E_R^{NC}, setting $Q^{NC} = Q$ does not influence the model's behavior regarding in-sample fit and out-of-sample predictions—after all, the primary goal of this exercise is to compare the performance of the two quantitative models with and without strategic complementarity.

Using the same procedure outlined in Section 8, I identify $N_{min} = \$29.1$ billion and $\lambda_N = 118.4$ from payment volume data. The undetermined parameters (E_R^{NC}, E_O^{NC}, E_I^{NC}, E_D^{NC}, c^{NC}, ψ^{NC}, $\overline{\varrho^{NC}}$, $\overline{\vartheta^{NC}}$) are inferred using the method of moments. This entails setting the finite-sample-analogous expectations of the following moments to zero, effectively leveraging the data to align the model's predictions with observed GCF repo rate behaviors:

$$\mathbb{E}\begin{bmatrix} \epsilon_r^t \\ \epsilon_r^t R_D^t \\ \epsilon_r^t R_O^t \\ \epsilon_r^t T_I^t \\ \epsilon_r^t T_D^t \\ \epsilon_r^t S_i^t \\ \epsilon_r^t E_{month}^t \end{bmatrix} = \mathbf{0}.$$

An additional moment ties bank i's beliefs to ensure that $E_D^{NC} T_D$ adequately reflects the variability in the realized lending quantities S_j. This is operationalized by setting the covariance between the error term from $S_j^t - E_D^{NC} T_D^t$ and T_D^t itself to zero:

$$\text{COV}[(S_j^t - E_D^{NC} T_D^t), T_D^t] = \mathbb{E}\left[\left((S_j^t - \mathbb{E}\left[S_j^t\right]) - E_D^{NC}(T_D^t - \mathbb{E}\left[T_D^t\right])\right)(T_D^t - \mathbb{E}\left[T_D^t\right])\right] = 0.$$

Consequently, the estimation procedure employs 8 moments to estimate 8 parameters.

In the absence of strategic complementarity, the model does not utilize information from payment delays. Furthermore, the assumptions regarding bank i's beliefs about bank j's repo lending practices are streamlined into a single parameter, E_D^{NC}. Despite these simplifications, this model adeptly captures the dynamic interactions among all other observable variables, maintaining consistency with the comprehensive approach outlined in Section 8. The estimated parameters are detailed in Table 8 and the model's performance is illustrated in Fig. 7.

Table 8: Estimated parameters for the model without strategic complementarity

Parameters	Meaning	Estimates
c^{NC}	late payment cost	102.3 (bps)
ψ^{NC}	overdraft cost	124.1 (bps)
E_R^{NC}	early payment strategy of bank j	0.648
E_I^{NC}	Treasury issuance effect	0.637
E_D^{NC}	Bank i believes lending quantity for bank j is $E_D^{NC}T_D$	0.0878
E_O^{NC}	early payment from non-dealer banks is $E_O^{NC}R_O$	0.322
$\overline{\varrho^{NC}}$	month-end effect	13.4 (bp)
$\overline{\vartheta^{NC}}$	other factors	-5.75 (bp)
Q^{NC}	regulatory minimum	16.2 ($bn)
N_{min}	minimum total payment volume	29.1 ($bn)
$1/\lambda_N$	$\mathbb{E}[N_i - N_{min}]$	8.4 ($bn)

E Internet Appendix: Estimating repo lending quantity of large dealer banks

To estimate the daily repo lending quantity of the large U.S. dealer banks, I obtain daily transaction volume underlying the calculation of Secured Overnight Financing Rate (SOFR) and Tri-Party General Collateral Rate (TGCR) from the New York Fed. Let $Volume_t$ be the difference between the underlying volume of SOFR and the underlying volume of TGCR. By construction, $Volume_t$ contains the total reserves lent by large U.S. dealer banks (i.e., the total reserve-draining intermediation in Fig. 17) on day t. However, because dealers are borrowing from and lending to each other, $Volume_t$ double counts the total repo lending quantity of the large dealer banks; $Volume_t$ also contains liquidity provided by other lenders in the repo market. Therefore, $Volume_t$ overstates the reserves lent by large U.S. dealer banks on day t. To estimate the fraction of $Volume_t$ that comes from reserves lent by large dealer banks, I use form 10-Q and call reports.[50] of the largest U.S. repo-active banks and U.S. subsidiaries of foreign banks that are subject to the same U.S. regulations. I extract the daily-average net overnight repo lending quantity of each U.S. GSIB in quarter q from those reports. I then take the sum of these quantities over all U.S. GSIBs in quarter q to get $\overline{S}_q^{GSIB}$. I assume $\overline{S}_q^{GSIB}$ approximates the daily average of total repo lending quantity of all large dealer banks in quarter q. Let $Volume_q$ be the daily average of $Volume_t$ in quarter q. I assume the ratio $\varphi_q = \overline{S}_q^{GSIB}/Volume_q$ is constant every day in quarter q. Thus, on day t in quarter q, the total repo lending quantity can be approximately calculated as $S_t = \varphi_q Volume_t$.

F Internet Appendix: Proofs

This appendix contains proofs.

F.1 Proof of Lemma 1

Consider bank i's decision problem. Conditional on the realizations of N_i and L_i, when $N_i \leq L_i$, the optimal strategy is $a_i = N_i$. Thus, for the remainder, I only consider the case in which $N_i > L_i$. The equilibrium strategy a_i of bank i is permitted to be a mixed strategy conditional on N_i and L_i. Any such mixed strategy a_i can be represented in the form $A(N_i, L_i, \epsilon_i)$ for some measurable, $A : [N_{min}, \infty) \times \mathbb{R} \times [0, 1] \to \mathbb{R}$ and some uniform random variable ϵ_i independent of $\{N_i, L_i, N_j, L_j, a_j\}$. Letting $\mathcal{A}$ denote the space of mixed payment strategies of this form, and given the payment strategy a_j of bank j, bank i solves

$$\operatorname*{ess\,inf}_{A \in \mathcal{A}} \quad \mathbb{E}[\psi(A(N_i, L_i, \epsilon_i) - L_i - a_j)^+ + c(N_i - A(N_i, L_i, \epsilon_i))^+ \mid N_i, L_i].$$

subject to

$$A(N_i, L_i, \epsilon_i) \geq 0 \qquad \text{almost surely}$$
$$A(N_i, L_i, \epsilon_i) \leq N_i \qquad \text{almost surely}.$$

An optimal A must satisfy $A(N_i, L_i, \epsilon) \geq L_i$ almost surely for all N_i, since $a_j \geq 0$ almost surely. In other words, any strategy A of bank i satisfying $A(N_i, L_i, \epsilon_i) < L_i$ with positive probability is dominated. Since bank j faces the same problem, bank i can correctly infer that $a_j \geq L_j$ by eliminating the dominated strategies of bank j.

Let $z = A(N_i, L_i, \epsilon_i) - L_i$. We have shown that $z \geq 0$ almost surely. The problem of

[50] The set of banks I use include JP Morgan Chase, Bank of America, Goldman Sachs, Morgan Stanley, Citi Bank, Wells Fargo Bank, PNC Bank, Deutsche Bank Trust Company Americas, HSBC Bank USA, and State Street.

bank i can be expressed as

$$\operatorname*{ess\,inf}_{z} \quad \mathbb{E}\Big[\mathbb{E}\big[\psi(z-a_j)^+ \;\big|\; z\big]\Big] + c(N_i - (\mathbb{E}[z] + L_i)) \tag{21}$$

subject to

$$z \geq (-L_i)^+ \qquad \text{almost surely} \tag{22}$$

$$z + L_i \leq N_i \qquad \text{almost surely,} \tag{23}$$

First we observe that when constraint (22) binds, $z = (-L_i)^+$ and $a_i = L_i^+$. Likewise, when constraint (22) binds, $z = N_i - L_i$ and $a_i = N_i$.

Let us consider the case when neither constraint (22) nor constraint (23) binds. Since $(z - a_j)^+$ is increasing and convex in z for all realizations of a_j, the mapping from real x to $\mathbb{E}\big[\psi(x-a_j)^+\big]$ is also convex. Thus, by Jensen's inequality, replacing z by $\mathbb{E}[z]$ weakly decreases the objective function (21). This implies that conditional on L_i, N_i, if bank i optimally chooses some z with support $\mathcal{I}_z \subset [-L_i, N_i - L_i]$, then for any υ_1 and υ_2 in $\mathcal{I}_z$ and any $\upsilon' \in [-L_i, N_i - L_i]$,

$$\mathbb{E}\Big[\psi(\upsilon_1 - a_j)^+\Big] + c(N_i - (\upsilon_1 + L_i)) = \mathbb{E}\Big[\psi(\upsilon_2 - a_j)^+\Big] + c(N_i - (\upsilon_2 + L_i)) \tag{24}$$

$$\mathbb{E}\Big[\psi(\upsilon_1 - a_j)^+\Big] + c(N_i - (\upsilon_1 + L_i)) \geq \mathbb{E}\Big[\psi(\upsilon' - a_j)^+\Big] + c(N_i - (\upsilon' + L_i)). \tag{25}$$

The first order condition for optimality in problem (10) implies that for any $\upsilon \in \mathcal{I}_z$,

$$\mathbb{P}(a_j \leq \upsilon) \geq \frac{c}{\psi}.$$

If there does not exist υ such that $\mathbb{P}(a_j \leq \upsilon) = c/\psi$, then $\mathcal{I}_z$ must be a singleton, and any optimal z is a constant function. We summarize the above arguments in the following lemma.

Lemma 9. *Suppose bank j choose any strategy a_j, the best response actions of bank i are of the form*

$$a_i = \min((L_i + \upsilon_i)^+, N_i),$$

where υ_i is some non negative random variable with support $\mathcal{I}_z$. For any $\upsilon \in \mathcal{I}_z$,

$$\mathbb{P}(a_j \leq \upsilon) \geq \frac{c}{\psi}.$$

Moreover, if there does not exist $\upsilon^\star$ such that $\mathbb{P}(a_j \leq \upsilon^\star) = c/\psi$, then $\mathcal{I}_z = \{\upsilon^*\}$ is a singleton and $\upsilon^* = \inf\{\vartheta \geq 0, \mathbb{P}(a_j \leq \vartheta) > \frac{c}{\psi}\}$. If there is an $\upsilon^\star$ such that $\mathbb{P}(a_j \leq \upsilon^\star) = c/\psi$, then for any $\upsilon \in \mathcal{I}_z$, $\mathbb{P}(a_j \leq \upsilon) = c/\psi$.

We first the characterize the class of pure strategy equilibrium.

Lemma 10. *There exists an equilibrium in the class of equilibria in which $\mathcal{I}_z$ is a singleton. In this equilibrium, bank i chooses the payment $a_i^* = \min((L_i + \alpha_i)^+, N_i)$, and bank j chooses payment*

$a_j^* = \min((L_j + \alpha_j)^+, N_j)$, *where* α_i *and* α_j *solve*

$$\alpha_i = \inf\left\{\vartheta \geq 0: \quad \mathbb{P}(\min((L_j + \alpha_j)^+, N_j) \leq \vartheta) \geq \frac{c}{\psi}\right\}$$
$$\alpha_j = \inf\left\{\vartheta \geq 0: \quad \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \vartheta) \geq \frac{c}{\psi}\right\}.$$

In particular, when L_i *and* L_j *have the same distribution, then* $\alpha_i = \alpha_j = \alpha$ *where*

$$\alpha = \inf\left\{\vartheta \geq 0: \ \mathbb{P}(N_j \leq \vartheta) \geq \frac{\frac{c}{\psi} - \mathbb{P}(L_j \leq 0)}{1 - \mathbb{P}(L_j \leq 0)}\right\}.$$

Proof. If $\mathcal{I}_z = \{\alpha_i\}$ is a singleton, taking constraints (22) and (23) into account, $a_i^* = \min((L_i + \alpha_i)^+, N_i)$. Given a_i^*, a similar analysis shows that $a_j^* = \min((L_j + \alpha_j)^+, N_j)$. Under the conditions of Lemma 1, N_i and N_j have the same distribution. When L_i and L_j also have the same distribution, then if $0 \leq \alpha_i < \alpha_j$,

$$\frac{c}{\psi} \leq \mathbb{P}(\min((L_j + \alpha_j)^+, N_j) \leq \alpha_i) \leq \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \alpha_i)$$
$$< \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \alpha_j) = \frac{c}{\psi}.$$

This is a contradiction. The last inequality follows from that F_N strictly increases on the interior of its support. Hence $\alpha_i \geq \alpha_j$. A symmetric argument shows that it must be that $\alpha_i = \alpha_j$. Denote this common value by α. By Lemma 9,

$$\alpha = \inf\left\{\vartheta \geq 0: \quad \mathbb{P}(\min((L_j + \alpha)^+, N_j) \leq \vartheta) \geq \frac{c}{\psi}\right\}.$$

When $\mathbb{P}(L_j \leq 0) \neq \frac{c}{\psi}$, the solution is unique. Since $N_i, N_j \in \mathbb{R}_{++}$,

$$\begin{aligned}\mathbb{P}(\min((L_j + \alpha)^+, N_j) \leq \alpha) &= \mathbb{P}(\min(L_j + \alpha, N_j) \leq \alpha) \\ &= \mathbb{P}(L_j + \alpha \leq \alpha) + \mathbb{P}(N_j \leq \alpha)(1 - \mathbb{P}(L_j + \alpha \leq \alpha)).\end{aligned}$$

This shows that α is determined by

$$\alpha = \inf\left\{\vartheta \geq 0: \quad \mathbb{P}(N_j \leq \vartheta) \geq \frac{\frac{c}{\psi} - \mathbb{P}(L_j \leq 0)}{1 - \mathbb{P}(L_j \leq 0)}\right\}.$$

By Lemma 9, when L_i and L_j have different distributions, α_i and α_j solve

$$\begin{aligned}\alpha_i &= \inf\left\{\vartheta \geq 0: \quad \mathbb{P}(\min((L_j + \alpha_j)^+, N_j) \leq \vartheta) \geq \frac{c}{\psi}\right\} \\ \alpha_j &= \inf\left\{\vartheta \geq 0: \quad \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \vartheta) \geq \frac{c}{\psi}\right\}.\end{aligned} \tag{26}$$

Clearly any possible solutions α_i, α_j is less than $F_N^{-1}(\frac{c}{\psi})$. Consider a map $\mathcal{T}^\alpha : [0, F_N^{-1}(\frac{c}{\psi})]^2 \to [0, F_N^{-1}(\frac{c}{\psi})]^2$, where $\mathcal{T}^\alpha(x, y) = (\mathfrak{a}_i, \mathfrak{a}_j)$ such that

$$\mathfrak{a}_i = \inf \left\{ \vartheta \geq 0 : \quad \mathbb{P}(\min((L_j + y)^+, N_j) \leq \vartheta) \geq \frac{c}{\psi} \right\}$$
$$\mathfrak{a}_j = \inf \left\{ \vartheta \geq 0 : \quad \mathbb{P}(\min((L_i + x)^+, N_i) \leq \vartheta) \geq \frac{c}{\psi} \right\}.$$

It is easy to check that $\mathcal{T}^\alpha$ is continuous. By Schauder fixed-point theorem there is at least one fixed point of $\mathcal{T}^\alpha$. □

Next, I rule out the possibility of mixed equilibria. Suppose that there is a mixed strategy equilibrium such that $a_i = \min((L_i + z_i)^+, N_i)$ and $a_j = \min((L_j + z_j)^+, N_j)$. Let $\mathcal{I}_z^i$ denote the support of z_i. Since $\mathcal{I}_z^i$ and $\mathcal{I}_z^j$ are bounded, let

$$\underline{\upsilon}_i \stackrel{def}{=} \inf\ \mathcal{I}_z^i, \qquad \underline{\upsilon}_j \stackrel{def}{=} \inf\ \mathcal{I}_z^j,$$
$$\overline{\upsilon}_i \stackrel{def}{=} \sup\ \mathcal{I}_z^i, \qquad \overline{\upsilon}_j \stackrel{def}{=} \sup\ \mathcal{I}_z^j.$$

At least one of $\mathcal{I}_z^i$ and $\mathcal{I}_z^j$ must have more than one element, for otherwise it is a pure strategy equilibrium. Say $\mathcal{I}_z^i$ has at least two elements, then $\overline{\upsilon}_i > \underline{\upsilon}_i$ and for any $\upsilon \in [\underline{\upsilon}_i, \overline{\upsilon}_i)$,

$$\mathbb{P}(a_j \leq \upsilon) = \frac{c}{\psi}. \tag{27}$$

Lemma 11. *There is no mixed strategy equilibrium.*

Proof. Suppose that there is a mixed strategy equilibrium. If $\underline{\upsilon}_i \geq N_{min}$ Pick any $N_{min} \leq \upsilon_i' < \upsilon_i''$ in $\mathcal{I}_z^i$. Since F_N strictly increases,

$$\mathbb{P}(N_j \leq \upsilon_i'') > \mathbb{P}(N_j \leq \upsilon_i'),$$

so if it is the case that

$$\mathbb{P}(a_j \leq \upsilon_i'') = \mathbb{P}(\min((L_j + z_j)^+, N_j) \leq \upsilon_i'') = \mathbb{P}(\min((L_j + z_j)^+, N_j) \leq \upsilon_i') = \mathbb{P}(a_j \leq \upsilon_i'),$$

then it must be that

$$\mathbb{P}((L_j + z_j)^+ \leq \upsilon_i' \ \big|\ N_j \in (\upsilon_i', \upsilon_i'']) = 1.$$

Thus, $\underline{\upsilon}_j \leq \max(\underline{\upsilon}_i, N_{min}) - \max(L_j) < \max(\underline{\upsilon}_i, N_{min})$ as $\mathbb{P}(L_j > 0) > 0$. If $\underline{\upsilon}_j = \overline{\upsilon}_j$ (i.e. z_j is degenerate), then $\mathbb{P}(\min((L_i + z_i)^+, N_i) \leq z_j) = 1 > \frac{c}{\psi}$. Thus, $\underline{\upsilon}_j < \overline{\upsilon}_j$. By a similar argument, $\underline{\upsilon}_i < \max(\underline{\upsilon}_j, N_{min})$. It follows that $\underline{\upsilon}_i < N_{min}$ and $\underline{\upsilon}_j < N_{min}$. Now if $\underline{\upsilon}_i \leq \underline{\upsilon}_j$,

$$\mathbb{P}(a_j \leq \underline{\upsilon}_i) = \mathbb{P}(L_j + z_j \leq \underline{\upsilon}_i) \leq \mathbb{P}(L_j \leq 0) < \frac{c}{\psi}.$$

This contradicts Eq. (27), so $\underline{\upsilon}_i > \underline{\upsilon}_j$, but a symmetric argument shows that $\underline{\upsilon}_i < \underline{\upsilon}_j$, contradiction. Thus, $\mathcal{I}_z^i$ and $\mathcal{I}_z^j$ must be singletons. □

It is easy to check that when $\mathbb{P}(L_j \leq 0) = c/\psi$, there are infinitely many equilibrium strategy profiles of the form $a_i = \min((L_i + \alpha)^+, N_i)$ for any $\alpha \in [0, N_{min})$.

Lemma 12. *Suppose L_i and L_j are arbitrarily distributed. There is a unique equilibrium to the payment game when*

1. *$\mathbb{P}(L_i \leq 0) > c/\psi$ and $\mathbb{P}(L_j \leq 0) > c/\psi$, in which case $\alpha_i = \alpha_j = 0$;*
2. *$\mathbb{P}(L_i \leq 0) < c/\psi$ and $\mathbb{P}(L_j \leq 0) < c/\psi$, in which case $\alpha_i > N_{min}$ and $\alpha_j > N_{min}$.*

Proof. Suppose that $\mathbb{P}(L_i \leq 0) > c/\psi$ and $\mathbb{P}(L_j \leq 0) > c/\psi$. If $\alpha_i \geq \alpha_j$ and $\alpha_i > 0$, then

$$\mathbb{P}(L_j + \alpha_j \leq \alpha_i) \geq \mathbb{P}(L_j \leq 0) > \frac{c}{\psi},$$

Bank i can be better off to deviate by making payment $\min(L_i^+, N_i)$. Thus, in this case both $\alpha_i = \alpha_j = 0$. Suppose that $\mathbb{P}(L_i \leq 0) < c/\psi$ and $\mathbb{P}(L_j \leq 0) < c/\psi$. Without loss of generality, assume $\alpha_i \geq \alpha_j$. If $\alpha_j < N_{min}$, then

$$\mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \alpha_j) = \mathbb{P}(L_i \leq \alpha_j - \alpha_i) < \frac{c}{\psi},$$

a contradiction to Lemma 9. Thus, $\alpha_j > N_{min}$ and $\alpha_i > N_{min}$. It can be easily checked that this case there can be only one solution to Eq. (26). □

F.2 Proof of Lemma 2

Let $F_{a_j}(o) = \mathbb{P}(a_j^* \leq o)$ where a_j^* is bank j's equilibrium strategy, then

$$\begin{aligned} V(L) = & \int_{n \in (L^+, (L+\alpha)^+]} \int_{o \in [0, n-L]} -\psi(n - L - o) dF_{a_j}(o) dF_N(n) + \\ & \int_{n \in ((L+\alpha)^+, \infty)} \int_{o \in [0, (L+\alpha)^+ - L]} -\psi\left((L+\alpha)^+ - L - o\right) dF_{a_j}(o) dF_N(n) + \\ & \int_{n \in ((L+\alpha)^+, \infty)} -c\left(n - (L+\alpha)^+\right) dF_N(n) \end{aligned}$$

Let ϵ be a small positive real number.

$$V(L-\epsilon) = \int_{n\in((L-\epsilon)^+,(L-\epsilon+\alpha)^+]} \int_{o\in[0,n-(L-\epsilon)]} -\psi(n-(L-\epsilon)-o)dF_{a_j}(o)dF_N(n)+$$
$$\int_{n\in((L-\epsilon+\alpha)^+,\infty)} \int_{o\in[0,(L-\epsilon+\alpha)^+-(L-\epsilon)]} -\psi\left((L-\epsilon+\alpha)^+-(L-\epsilon)-o\right)dF_{a_j}(o)dF_N(n)+$$
$$\int_{n\in((L-\epsilon+\alpha)^+,\infty)} -c\left(n-(L-\epsilon+\alpha)^+\right)dF_N(n)$$

Direct calculation yields

$$\begin{aligned}
\Gamma(L) =& \lim_{\epsilon\to 0}\frac{V(L)-V(L-\epsilon)}{\epsilon}\\
=& \lim_{\epsilon\to 0}\int_{n\in[L^+,(L-\epsilon+\alpha)^+]}\int_{o\in[0,n-L]}\psi\, dF_{a_j}(o)dF_N(n)\\
&+\mathbb{P}(a_j\le(L+\alpha)^+-L)\mathbb{P}(N=(L+\alpha)^+)\psi\lim_{\epsilon\to 0}\frac{(L-\epsilon+\alpha)^+-(L+\alpha)^++\epsilon}{\epsilon}\\
&+\lim_{\epsilon\to 0}c\mathbb{P}(N=(L+\alpha)^+)\frac{((L+\alpha)^+-(L-\epsilon+\alpha)^+)}{\epsilon}\\
&+\lim_{\epsilon\to 0}\int_{n\in((L+\alpha)^+,\infty)}\frac{\begin{array}{r}\psi\left((L-\epsilon+\alpha)^+-(L+\alpha)^++\epsilon\right)F_{a_j}((L+\alpha)^+-L)\\+c\left((L+\alpha)^+-(L-\epsilon+\alpha)^+\right)\end{array}}{\epsilon}dF_N(n).
\end{aligned}$$

Which can be further simplified to the desired results.

F.3 Proofs of results in Section 4.2

In Section 4.2 I study the competitive equilibrium in the wholesale funding markets, and in Appendix D.1 I study monopolistic pricing. It turns out that the proofs for the competitive case are the simpler version of the proofs for the monopolistic pricing case. From Appendix F.4 to Appendix G.1, I state the detailed proofs for all results in Appendix D.1. To save space, I will only briefly state how to change those proofs to prove the results stated in Section 4.2 here.

I first state an important lemma that links the endogenous distribution of L_i and L_j with the exogenous state variables, given the payment subgame strategy profile.

Lemma 13. *Fix the payment subgame equilibrium strategy profile* $\{a_i^* = \min((L_i+\alpha_i)^+, N_i)\}$.

For any $o \in \mathbb{R}$, and reserve balances after trade $L_j = R_j - Q - S_j$ possible in equilibrium,

$$L_j < \ o \iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} < \ o$$
$$L_j \geq \ o \iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} \geq \ o.$$

Let $\Gamma_j^+(o, \alpha_j, \alpha_i) = \lim_{x \downarrow o} \Gamma_j(x, \alpha_j, \alpha_i)$, then

$$\mathbb{P}(L_j \leq \ o) = \mathbb{P}(R_j - Q - D_j + \frac{\Gamma_j^+(o, \alpha_j, \alpha_i)}{\xi} \leq \ o) = \mathbb{E}\left[F_R\left(D_j + Q - \frac{\Gamma_j^+(o, \alpha_j, \alpha_i)}{\xi}\right)\right].$$

Proof. The proof is identical to the proof of Lemma 4 after we set $\lambda^{-1} = 0$ in the proof of Lemma 4. □

Comparing Lemma 13 and Lemma 4, we see same results except that in the competitive case, we omit the term λ^{-1}. The information rent λ^{-1} for borrower is zero under competitive pricing, but is positive under monopolistic screening, as discussed in Appendix D.1. In fact: the case of competitive pricing where bank i and borrower i acts as price takers in Section 4.2 is a simpler version of case of monopolistic screening.

Proof of Theorems 1 to 3. Rewrite λ^{-1} as zero in Appendices F.6 to F.8. The rest of the proofs are exactly the same. □

For the proofs of Theorems 4 and 5, I abuse notations and write $\Gamma(y, \alpha, \alpha)$ as $\Gamma(y\ ;\ m)$ to denote the marginal value of liquidity at y when the liquidity stress index is m under some macroeconomic condition $\mathcal{M}_C^m$. Under macroeconomic condition $\mathcal{M}_C^0$, there could be multiple equilibria. For the following analysis, we select the equilibrium with the $\alpha = N_{min}$, because this equilibrium is the limiting equilibrium of a sequence of economies under no hoarding condition such that $m \uparrow 0$.

Proof of Theorem 4. In the following proof, we fix the parameters c, Q, and ψ from one set of macroeconomic conditions with liquidity stress index 0, denoted by $\mathcal{M}_C^0$. We will consider a sequence of $\mathcal{M}_C^{\epsilon^m}$ with $\epsilon^m \downarrow 0$. (Given that $\mathcal{M}_C^0$ and $\mathcal{M}_C^{\epsilon^m}$ are close with respect to the liquidity stress index, differences in their constants and functions become negligible for small ϵ^m. Therefore, whether under hoarding equilibrium or no hoarding equilibrium, the effects of those differences on equilibrium funding rates diminish. The distinctions between these sets are primarily significant for their influence on the liquidity stress index, which subsequently determines different equilibrium states, as illustrated below.) It suffices to show that for all small $\epsilon^m > 0$, there is some $\delta_0(\zeta, s) \geq 0$ and constant $\mathcal{O}^1 \geq 0$ such that

$$\Gamma(\zeta - Q - s\ ;\ \epsilon^m) - \Gamma(\zeta - Q - s\ ;\ 0) > \delta_0(\zeta, s) - \mathcal{O}^1 \epsilon^m.$$

In addition, whenever $\zeta - Q - s \neq 0$, $\delta_0(\zeta, s) > 0$. Recall that under no hoarding condition, $a_j^* = \min((L_j + \alpha)^+, N_j)$ for some $\alpha \geq N_{min}$. Under no hoarding condition, $a_j^* = \min(L_j^+, N_j)$. Note that $\Gamma(y\,;\,0) \geq \Gamma(y\,;\,\epsilon)$ for any ϵ. We discuss $\Gamma(y\,;\,\epsilon) - \Gamma(y\,;\,0)$ case by case. First, when $y \geq 0$, by directly calculation the assumption that two macroeconomic conditions are close with respect to liquidity stress index, there is some constant $\mathcal{O}^3 > 0$ such that

$$\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) \geq \int_y^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j + \alpha)^+, N_j\} \leq n - y)\, d\, F_N(n) - \mathcal{O}^3\epsilon^m.$$

When $y \geq N_{min}$,

$$\begin{aligned}
&\int_y^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j + \alpha)^+, N_j\} \leq n - y)\, d\, F_N(n) \\
&= \int_{y^+}^{y+\frac{N_{min}}{2}} c - \psi\mathbb{P}(\min\{L_j + \alpha, N_j\} \leq n - y)\, d\, F_N(n) + \int_{y+\frac{N_{min}}{2}}^{y+\alpha} c - \psi\mathbb{P}(\min\{L_j + \alpha, N_j\} \leq n - y)\, d\, F_N(n) \\
&\geq \int_{y^+}^{y+\frac{N_{min}}{2}} c - \psi\mathbb{P}(L_j + \alpha \leq \frac{N_{min}}{2})\, d\, F_N(n) + \int_{y+\frac{N_{min}}{2}}^{y+\alpha} c - \psi\mathbb{P}(\min\{L_j + \alpha, N_j\} \leq n - y)\, d\, F_N(n).
\end{aligned}$$

Since $c - \psi\mathbb{P}(\min\{L_j + \alpha, N_j\} \leq \alpha) = 0$ and $\mathbb{P}(\min\{L_j + \alpha, N_j\} \leq \vartheta)$ strictly increases in ϑ, there is some $\delta_1(\zeta, y)$ such that

$$\int_{y^+}^{y+\frac{N_{min}}{2}} c - \psi\mathbb{P}(L_j + \alpha \leq \frac{N_{min}}{2})\, d\, F_N(n) > \delta_1(\zeta, y) > 0,$$

and

$$\int_{y+\frac{N_{min}}{2}}^{y+\alpha} c - \psi\mathbb{P}(\min\{L_j + \alpha, N_j\} \leq n - y)\, d\, F_N(n) \geq 0.$$

Thus,

$$\int_y^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j + \alpha)^+, N_j\} \leq n - y)\, d\, F_N(n) > \delta_1(\zeta, y) > 0,$$

Hence, when $y \geq N_{min}$,

$$\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) > \delta_1(\zeta, y) - \mathcal{O}^3\epsilon^m.$$

When $N_{min} > y > 0$,

$$\begin{aligned}\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) &\geq \int_{N_{min}}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n - y)\, d\,F_N(n) - \mathcal{O}^3\epsilon^m \\ &> \int_{N_{min}}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n - N_{min})\, d\,F_N(n) - \mathcal{O}^3\epsilon^m.\end{aligned}$$

When $y = 0$, clearly $\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) \geq -\mathcal{O}^3\epsilon^m$. When $0 > y \geq -\alpha$, $\exists \mathcal{O}^4 > 0$ such that

$$\begin{aligned}&\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) \\ &\geq \psi\,\mathbb{P}^{\epsilon^m}(\min(L_j^+, N_j) \leq -y) - \psi\int_0^{y+\alpha} \mathbb{P}(a_j^* \leq n - y) d\,F_N(n) - c\,\mathbb{P}(N_i > y+\alpha) - \mathcal{O}^4\epsilon^m \\ &> \delta_2(\zeta, y) + c - \psi\int_0^{y+\alpha} \mathbb{P}(a_j^* \leq n - y\,;\,T+\epsilon) d\,F_N(n) - c\,\mathbb{P}(N_i > y+\alpha) - \mathcal{O}^4\epsilon^m > \delta_2(\zeta, y) - \mathcal{O}^4\epsilon^m.\end{aligned}$$

where the second inequality derives from the fact that under liquidity hoarding condition, $\psi\,\mathbb{P}^{\epsilon^m}(L_j \leq 0) > c$. Finally, when $y < -\alpha$, then

$$\begin{aligned}&\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) = \psi\,\mathbb{P}^{\epsilon^m}(\min(L_j^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min((L_j+\alpha)^+, N_j) \leq -y) \\ &> \psi\,\mathbb{P}^{\epsilon^m}(\min(L_j^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min(L_j^+, N_j) \leq -y) \\ &\quad + \psi\left(\mathbb{P}(\min((L_j+\alpha)^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min(L_j^+, N_j) \leq -y)\right) \\ &> \psi\left(\mathbb{P}(\min((L_j+\alpha)^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min((L_j)^+, N_j) \leq -y)\right) > \delta_3(\zeta, y)\end{aligned}$$

for some $\delta_3(\zeta, y) > 0$ regardless of ϵ. Let $\delta_0(\zeta, \zeta - Q - y)$ be the corresponding $\delta_i(\zeta, y)$ in each case corresponding to different y. This finishes the proof. □

Proof of Theorem 5. Based on the proof above, for all small ϵ^m, there is some $\delta_0(\zeta, \mathcal{S}^*(\zeta, \mathscr{D}, 0)) \geq 0$ and constant $\mathcal{O}^1 \geq 0$ such that $\Gamma(\zeta - Q - \mathcal{S}^*(\zeta, \mathscr{D}, 0)\,;\,\epsilon^m) - \Gamma(\zeta - Q - \mathcal{S}^*(\zeta, \mathscr{D}, 0)\,;\,0) > \delta_0(\zeta, \mathcal{S}^*(\zeta, \mathscr{D}, \epsilon^m)) - \mathcal{O}^1\epsilon^m$ and $\delta_0(\zeta, \mathcal{S}^*(\zeta, \mathscr{D}, 0)) > 0$. For ϵ^m small enough, $\exists \delta_1(\zeta, \mathscr{D}) > 0$ such that

$$\Gamma_i(\zeta - Q - \mathcal{S}^*(\zeta, \mathscr{D}, 0)\,;\,\epsilon^m) + \xi\mathcal{S}^*(\zeta, \mathscr{D}, 0) > \Gamma_i(\zeta - Q - \mathcal{S}^*(\zeta, \mathscr{D}, 0)\,;\,0) + \xi\mathcal{S}^*(\zeta, \mathscr{D}, 0) + \delta_1(\zeta, \mathscr{D}).$$

From Eq. (4), this implies that

$$\mathcal{S}^*(\zeta, \mathscr{D}, 0) - \mathcal{S}^*(\zeta, \mathscr{D}, \epsilon^m) > \delta(\zeta, \mathscr{D})$$

for some $\delta(\zeta, \mathscr{D}) > 0$. Since $r^*(\zeta, \mathscr{D}, \epsilon^m) = \xi(\mathscr{D} - \mathcal{S}^*(\zeta, \mathscr{D}, \epsilon^m))$, the result follows. □

F.4 Proof of Lemma 3

First consider the following relaxed version of the problem for bank i. Suppose that bank i designs a mechanism for selecting the transaction quantity S_i and repo rates r_i. By the revelation principle (Myerson, 1986), we can focus on direct mechanisms without loss of generality. I will characterize the optimal direct mechanism and show that the optimal direct mechanism can be implemented by a supply schedule.

The type space of a repo borrower is $[D_{min}, \infty)$. A direct mechanism consists of functions $\mathcal{Q} : [D_{min}, \infty) \to \mathbb{R}$ and $T : [D_{min}, \infty) \to \mathbb{R}$. The direct mechanism design problem of bank i is

$$\sup_{\mathcal{Q},T} \quad \mathbb{E}[V(R_i - Q - \mathcal{Q}(D)) + T(D)],$$

subject to the incentive-compatibility (IC) constraint

$$-\frac{\xi}{2}\left((D - \mathcal{Q}(D))^+\right)^2 - T(D) \geq -\frac{\xi}{2}\left((D - \mathcal{Q}(\theta))^+\right)^2 - T(\theta), \qquad (D, \theta) \in [D_{min}, \infty)^2,$$

and the individually rational (IR) constraint

$$-\frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - T(D) \geq -\frac{\xi}{2}D^2.$$

Lemma 14. *Under any optimal contract, $\mathcal{Q}(D) \leq D$ for all D in $[D_{min}, \infty)$.*

Proof. When $\mathcal{Q}(D) > D$ for some D, bank i can set $\mathcal{Q}(D) = D$ without changing the cost of the repo borrower, thus respecting the IC constraints of all types. However, this would weakly increase the payoff of bank i. □

By Lemma 14, we can focus without loss of generality on a mechanism $(\mathcal{Q}, T)$ satisfying $\mathcal{Q}(D) \leq D$. In the following lemma, I characterize the set of allocations and payment strategies that satisfy the IC constraint. This simplifies the subsequent analysis.

Lemma 15. *For any mechanism $(\mathcal{Q}, T)$, any incentive compatible allocation rule $\mathcal{Q}$ is weakly increasing with D.*

Proof. Consider two borrower types D and D' with $D > D'$. Incentive compatibility requires that

$$-\frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - T(D) \geq -\frac{\xi}{2}\Big((D - \mathcal{Q}(D'))^+\Big)^2 - T(D')$$

and that

$$-\frac{\xi}{2}\Big((D'-\mathcal{Q}(D'))^+\Big)^2 - T(D') \geq -\frac{\xi}{2}\Big((D'-\mathcal{Q}(D))^+\Big)^2 - T(D).$$

Adding these inequalities, we see that

$$-\frac{\xi}{2}\Big((D-\mathcal{Q}(D))^+\Big)^2 - \frac{\xi}{2}\Big((D'-\mathcal{Q}(D'))^+\Big)^2 \geq -\frac{\xi}{2}\Big((D'-\mathcal{Q}(D))^+\Big)^2 - \frac{\xi}{2}\Big((D-\mathcal{Q}(D'))^+\Big)^2.$$

Rearranging and invoking Lemma 14, we get

$$\Big((D'-\mathcal{Q}(D))^+ + (D-\mathcal{Q}(D))\Big)\Big(D-\mathcal{Q}(D)-(D'-\mathcal{Q}(D))^+\Big) \leq \\ (D'+D-2Q(D'))\,(D-D')\,. \tag{28}$$

If $\mathcal{Q}(D) < \mathcal{Q}(D')$, then $\mathcal{Q}(D) < D'$ and

$$\Big((D'-\mathcal{Q}(D))^+ + (D-\mathcal{Q}(D))\Big)\Big(D-\mathcal{Q}(D)-(D'-\mathcal{Q}(D))^+\Big) = \\ \Big(D'-\mathcal{Q}(D)+D-\mathcal{Q}(D)\Big)(D-D') > (D'+D-2Q(D'))\,(D-D')\,,$$

contradicting inequality (28). Thus, it is necessary that $\mathcal{Q}(D) \geq \mathcal{Q}(D')$. □

Let $u:\mathbb{R}\to\mathbb{R}$ be the value function of a repo borrower under a given mechanism $(\mathcal{Q}, T)$, in that

$$u(D) \stackrel{def}{=} \sup_{\theta\in[D_{min},\infty)} \quad -\frac{\xi}{2}\Big((D-\mathcal{Q}(\theta))^+\Big)^2 - T(\theta).$$

Lemma 16. *Any incentive compatible truthful mechanism* $(\mathcal{Q}, T)$ *must satisfy*

$$u(D) = -\frac{\xi}{2}\Big((D-\mathcal{Q}(D))^+\Big)^2 - T(D) = u(D_{min}) - \int_{D_{min}}^{D} \xi(x-\mathcal{Q}(x))\,dx.$$

Proof. By Lemma 14, we can focus on an IC mechanism $(\mathcal{Q}, T)$ with $\mathcal{Q}(D) \leq D$. We first show that u is absolutely continuous on any bounded interval. Let $M > 0$. We claim that $u(\cdot)$ is absolutely continuous on $(-M, M)$. Indeed, a truthful incentive compatible mechanism must satisfy

$$u(D) = -\frac{\xi}{2}\Big((D-\mathcal{Q}(D))^+\Big)^2 - T(D) \geq -\frac{\xi}{2}\Big((D-\mathcal{Q}(\theta))^+\Big)^2 - T(\theta)$$
$$u(\theta) = -\frac{\xi}{2}\Big((\theta-\mathcal{Q}(\theta))^+\Big)^2 - T(\theta) \geq -\frac{\xi}{2}\Big((\theta-\mathcal{Q}(D))^+\Big)^2 - T(D),$$

whenever $-M < \theta < D < M$. Thus,

$$u(D) - u(\theta) \leq -\frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 + \frac{\xi}{2}\Big((\theta - \mathcal{Q}(D))^+\Big)^2$$
$$= \frac{\xi}{2}\Big((\theta - \mathcal{Q}(D))^+ - (D - \mathcal{Q}(D))\Big)\Big((\theta - \mathcal{Q}(D))^+ + (D - \mathcal{Q}(D))\Big) \leq 0$$

and

$$u(D) - u(\theta) \geq -\frac{\xi}{2}\Big((D - \mathcal{Q}(\theta))^+\Big)^2 + \frac{\xi}{2}\Big((\theta - \mathcal{Q}(\theta))^+\Big)^2 = \frac{\xi}{2}\Big(\theta - D\Big)\Big(\theta + D - 2\mathcal{Q}(D)\Big).$$

Thus,

$$\left|\frac{u(D) - u(\theta)}{D - \theta}\right| \leq \frac{\xi}{2}\Big(\theta + D - 2\mathcal{Q}(D)\Big) \leq \frac{\xi}{2}\Big(2M - 2\mathcal{Q}(-M)\Big),$$

where the last inequality holds due to the monotonicity of $\mathcal{Q}$ shown in Lemma 15. The above inequality holds true whenever $-M < \theta < D < M$.

Finally, since the above argument holds for arbitrary $M > 0$, the envelope theorem (Milgrom and Segal, 2002) implies that for all D, $u'(D) = -\xi(D - \mathcal{Q}(D))$. By the fundamental theorem of calculus,

$$-\frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - T(D) = u(D) = u(D_{min}) - \int_{D_{min}}^{D} \xi(x - \mathcal{Q}(x))\,dx.$$

□

Lemma 17. *A direct mechanism $(\mathcal{Q}, T)$ satisfying $\mathcal{Q}(D) \leq D$ for all D is incentive-compatible if and only if*

1. *$\mathcal{Q}(D)$ is weakly increasing in D.*
2. *$T(D) = -\frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min}) + \int_{D_{min}}^{D} \xi(x - S(x))\,dx$.*

Proof. We just need to show the "if" part. We need to show the IC condition

$$u(D) \geq -\frac{\xi}{2}\Big((D - \mathcal{Q}(\theta))^+\Big)^2 - T(\theta).$$

Substituting $T(D)$, the IC condition holds if and only if

$$-\int_{\theta}^{D} \xi(x - \mathcal{Q}(x))\,dx \geq -\frac{\xi}{2}\Big((D - \mathcal{Q}(\theta))^+\Big)^2 + \frac{\xi}{2}\Big((\theta - \mathcal{Q}(\theta))^+\Big)^2,$$

which holds since $\mathcal{Q}(D)$ is weakly increasing. □

We now turn to a bank's problem. The bank will take D as random. Since T is determined by $\mathcal{Q}$, we substitute $T(D)$ from Lemma 3 to get

$$\sup_{\mathcal{Q}} \mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min}) + \int_{D_{min}}^{D} \xi(x - \mathcal{Q}(x))\, dx\right], \quad (29)$$

subject to the condition that $\mathcal{Q}(D)$ is weakly increasing.

Using standard trick of integration by part as in Börgers (2015), we get

$$\begin{aligned}
&\mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min}) + \int_{D_{min}}^{D} \xi(x - \mathcal{Q}(x))\, dx\right] \\
&= \mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min})\right] + \int_{D_{min}}^{\infty}\int_{D_{min}}^{D} \xi(x - \mathcal{Q}(x))\, dx f_D(D) dD \\
&= \mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min})\right] + \int_{D_{min}}^{\infty}\int_{x}^{\infty} \xi(x - \mathcal{Q}(x)) f_D(D)\, dD dx \\
&= \mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min})\right] + \int_{D_{min}}^{\infty} \xi(x - \mathcal{Q}(x)) \frac{1 - F_D(x)}{f_D(x)} f_D(x)\, dx \\
&= \mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min})\right] + \int_{D_{min}}^{\infty} \xi(D - \mathcal{Q}(D)) \frac{1 - F_D(D)}{f_D(D)} f_D(D)\, dD \\
&= \mathbb{E}\left[V(R_0 - Q - \mathcal{Q}(D)) - \frac{\xi}{2}\Big((D - \mathcal{Q}(D))^+\Big)^2 - u(D_{min}) + \xi(D - \mathcal{Q}(D)) \frac{1 - F_D(D)}{f_D(D)}\right].
\end{aligned}$$

Thus, the solution to (29) must satisfy

$$\begin{aligned}
\mathcal{Q}(D_i) &= \inf\left\{s : \Gamma(R_i - Q - s) \geq \xi(D_i - s - \tfrac{1 - F_D(D_i)}{f_D(D_i)})\right\} \\
T(D_i) &= -\frac{\xi}{2}\Big(D - \mathcal{Q}(D)\Big)^2 + \frac{\xi}{2} {D_{min}}^2 + \int_{D_{min}}^{D_i} \xi(x - \mathcal{Q}(x))\, dx,
\end{aligned} \quad (30)$$

where F_D is the cumulative distribution function of D. The mapping from D to $D - \frac{1-F(D)}{f(D)}$ is strictly increasing by the assumed form of the density $f(\cdot)$. Thus, $\mathcal{Q}(\cdot)$ is weakly increasing. Finally, the IR condition is satisfied because

$$
\begin{aligned}
u(D) =& u(D_{min}) - \int_{D_{min}}^{D} \xi(x - \mathcal{Q}(x))\,dx = -\frac{\xi}{2}D_{min}^2 - \int_{D_{min}}^{D} \xi(x - \mathcal{Q}(x))\,dx \\
\geq & -\frac{\xi}{2}D_{min}^2 - \int_{D_{min}}^{D} \xi x\,dx = -\frac{\xi}{2}D^2.
\end{aligned}
$$

Lemma 18. *The repo game has a unique equilibrium. This unique equilibrium is fully separating. The outcome (r_i, S_i) is the same as that implied by the direct mechanism $(\mathcal{Q}, T)$ defined by Eq. (30).*

Proof. The direct mechanism can be implemented by posting a schedule. This is implied by the taxation principle (Mussa and Rosen, 1978). Indeed, by the IC constraint, if $\mathcal{Q}(D) = \mathcal{Q}(D')$, then $T(D) = T(D')$. The mapping $D \mapsto (\mathcal{Q}(D), T(D))$ generates a graph on $\mathbb{R}^2$ whose trace defines an associated supply schedule g. More specifically, let $\mathcal{Q}^{-1} : \mathcal{Q}([D_{min}, \infty)) \to \mathbb{R}$ be the inverse function of quantity defined by

$$
\mathcal{Q}^{-1}(t) = \inf \Big\{ \vartheta \in [D_{min}, \infty) : \ \mathcal{Q}(\vartheta) = t \Big\}.
$$

Then the supply schedule g is defined on $\mathcal{Q}([D_{min}, \infty))$ by

$$
g(s) = \frac{T(\mathcal{Q}^{-1}(s))}{s}.
$$

When $s \notin \mathcal{Q}([D_{min}, \infty))$, define $g(s) = \infty$. Since the direct mechanism $(\mathcal{Q}, T)$ maximizes the payoff of bank i, the associated supply schedule g also maximizes the payoff of bank i. This direct mechanism is dominant implementable. Thus, the repo game has a unique equilibrium. This unique equilibrium is separating. □

F.5 Proof of Lemma 4

Fix any $o \in \mathbb{R}$. By Lemma 2 $\Gamma_j(y, \alpha_j, \alpha_i)$ is weakly decreasing in y. By Lemma 3, $\Gamma_j(R_j - Q - S_j, \alpha_j, \alpha_i) \geq \xi(D_j - S_j - \lambda^{-1})$ in equilibrium. Hence, when $L_j = R_j - Q - S_j \geq o$,

$$
\begin{aligned}
\Gamma_j(R_j - Q - S_j, \alpha_j, \alpha_i) \leq \Gamma_j(o, \alpha_j, \alpha_i) &\Rightarrow D_j - S_j - \lambda^{-1} \leq \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} \Rightarrow \\
R_j - Q - S_j &\leq \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} - D_j + R_j - Q + \lambda^{-1} \Rightarrow \\
o \leq L_i &\leq \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} - D_j + R_j - Q + \lambda^{-1}.
\end{aligned}
$$

Claim 1. Suppose S_j is the equilibrium trading quantity. Then for any $S' < S_j$, $\Gamma_j(R_j - Q - S', \alpha_j, \alpha_i) \leq \xi(D_j - S_j - \lambda^{-1})$.

Suppose $o \in \mathbb{R}$ such that $L_j = R_j - Q - S_j < o$. Let $S_o = R_j - Q - o < S_j = R_j - Q - L_j$. Claim 1 implies that $\Gamma_j(o, \alpha_j, \alpha_i) = \Gamma_j(R_j - Q - S_o, \alpha_j, \alpha_i) \leq \xi(D_j - S_j - \lambda^{-1})$. Thus,

$$R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} \leq R_j - Q - S_j = L_j < o.$$

Thus,

$$L_j = R_j - Q - S_j \geq o \iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} \geq o$$
$$L_j = R_j - Q - S_j < o \iff R_j - Q - D_j + \frac{\Gamma_j(o, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} < o.$$

Finally, since probability measure is continuous from above, $\mathbb{P}(L_j \leq o) = \mathbb{P}(R_j - Q - D_j + \frac{\Gamma_j^+(o,\alpha_j,\alpha_i)}{\xi} + \lambda^{-1} \leq o)$.

Proof of Claim 1. By definition, $S_j = \inf\left\{s : \Gamma_j(R_j - Q - s, \alpha_j, \alpha_i) \geq \xi(D_j - s - \lambda^{-1})\right\}$. Thus, $\Gamma_j(R_j - Q - S_j, \alpha_j, \alpha_i) + \xi S_j \geq \xi(D_j - \lambda^{-1})$. In addition, for any $S'' < S_j$,

$$\Gamma_j(R_j - Q - S'', \alpha_j, \alpha_i) + \xi S'' < \xi(D_j - \lambda^{-1}) \Rightarrow \lim_{s \uparrow S_j}(\Gamma_j(R_j - Q - s, \alpha_j, \alpha_i) + \xi s) \leq \xi(D_j - \lambda^{-1}) \Rightarrow$$
$$\lim_{s \uparrow S_j} \Gamma_j(R_j - Q - s, \alpha_j, \alpha_i) + \xi S_j \leq \xi(D_j - \lambda^{-1}).$$

Since $\Gamma_j(y, \alpha_j, \alpha_i)$ is weakly decreasing in y, for any $S' < S_j$, $\Gamma_j(R_j - Q - S', \alpha_j, \alpha_i) + \xi S_j \leq \xi(D_j - \lambda^{-1})$. □

F.6 Proof of Theorem 6

Suppose that liquidity hoarding condition Eq. (13) holds for $j \in \{1, 2\}$. We prove by contradiction that banks hoard liquidity in equilibrium at time 1. Assume that bank j does not hoard liquidity, i.e. bank j pays $\min(L_j^+ + z_j, N_j)$ at time 1 for some non-degenerate random variable $z_j \geq 0$. Let $i \in \{1, 2\} \setminus \{j\}$. Let $\mathcal{I}_j$ be the support of z_j. Trivially, $\mathcal{I}_j$ is bounded above. By Lemma 9, for any $\upsilon_j \in \mathcal{I}_j \cap (0, \infty)$, either $\mathbb{P}(a_i \leq \upsilon_j) = \frac{c}{\psi}$ or $\upsilon_j = \inf\{\vartheta \geq 0, \mathbb{P}(a_j \leq \vartheta) > \frac{c}{\psi}\}$. Let $V'_{j,-}(y) = \mathbb{E}[\Gamma_j^+(y, z_j, z_i)]$ be the marginal value of liquidity for bank j, where the expectation is taken over the random realizations of the mixed strategy profile. By Lemma 2, $V'_{j,-}(0) \leq c$. Then by Lemma 4 and Eq. (13),

$$\mathbb{P}(L_j \leq 0) = \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \leq -\frac{V'_{j,-}(0)}{\xi}\right) > \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \leq -\frac{c}{\xi}\right) > \frac{c}{\psi}.$$

If bank i does not hoard liquidity either, then $\mathbb{P}(L_i \leq 0) > \frac{c}{\psi}$ as well. However, as shown in Lemma 12, when $\mathbb{P}(L_i \leq 0) > \frac{c}{\psi}$ and $\mathbb{P}(L_j \leq 0) > \frac{c}{\psi}$, banks must both hoard liquidity, a contradiction. If bank i hoard liquidity but bank j does not hoard liquidity, then $\mathbb{P}(L_j \leq 0) > \frac{c}{\psi}$. Since bank j is best responding, it must still hold that for any $\upsilon_j \in \mathcal{I}_j \cap (0, \infty)$, either $\mathbb{P}(a_i \leq \upsilon_j) = \frac{c}{\psi}$, or $\mathbb{P}(a_i \leq \upsilon_j) > \frac{c}{\psi}$ and $\upsilon_j = \inf\{\vartheta \geq 0, \mathbb{P}(a_j \leq \vartheta) > \frac{c}{\psi}\}$.

Note that

$$\mathbb{P}(a_i \le \upsilon_j) = \mathbb{P}(\min((L_i)^+, N_i) \le \upsilon_j) \ge \mathbb{P}(L_i \le \upsilon_j) = \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le \upsilon_j - \frac{V'_{i,-}(\upsilon_j)}{\xi}\right).$$

From Lemma 2, $V'_{i,-}(\upsilon_j) \le c$ when $\upsilon_j > 0$. Thus,

$$\mathbb{P}(L_i \le \upsilon_j) = \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le \upsilon_j - \frac{V'_{i,-}(\upsilon_j)}{\xi}\right) \ge \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le -\frac{c}{\xi}\right) > \frac{c}{\psi}.$$

If $\mathbb{P}(a_i \le \upsilon_j) = \frac{c}{\psi}$, then $\frac{c}{\psi} > \frac{c}{\psi}$, a contradiction. If $\mathbb{P}(a_i \le \upsilon_j) > \frac{c}{\psi}$ and $\upsilon_j = \inf\{\vartheta \ge 0, \mathbb{P}(a_j \le \vartheta) > \frac{c}{\psi}\}$, then pick a very small $\epsilon > 0$. Since $F_N(\cdot)$ is strictly increasing, $\mathbb{P}(a_i \le \upsilon_j) \ge \mathbb{P}(a_i \le \upsilon_j - \epsilon) > \frac{c}{\psi}$, a contradiction.

F.7 Proof of Theorem 7

Suppose that the no hoarding condition Eq. (14) holds for $j \in \{1,2\}$. By Lemma 4, the assumption $\mathbb{P}(R_i - Q - D_{min} > 0) > 0$ and Lemma 1, no banks will not play mixed strategy in the payment subgame in equilibrium. Furthermore, bank i pays $a_i = \min((L_i + \alpha_i)^+, N_i)$ and bank j pays $a_j = \min((L_j + \alpha_j)^+, N_j)$ and

$$\alpha_i = \inf\{\vartheta \ge 0, \mathbb{P}(a_j \le \vartheta) \ge \frac{c}{\psi}\}, \qquad \alpha_j = \inf\{\vartheta \ge 0, \mathbb{P}(a_i \le \vartheta) \ge \frac{c}{\psi}\}.$$

Our first goal is to show that the real numbers $\alpha_i, \alpha_j \ge N_{min}$. First, by Lemma 2 and Definition 1,

$$\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i) = \begin{cases} \int\limits_{n \in ((\alpha_i - \alpha_j)^+, \alpha_i)} \psi \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \le n + \alpha_j - \alpha_i)\, dF_N(n) + \int\limits_{n \in [\alpha_i, \infty)} c dF_N(n), & \text{if } \alpha_i > 0; \\ \psi \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \le \alpha_j - \alpha_i), & \text{if } \alpha_i \le 0. \end{cases}$$

Thus, if $\alpha_i < N_{min}$, $\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i) \ge c$. If $\alpha_i \le \alpha_j$, then Eq. (14) implies that

$$\mathbb{P}(R_j - Q - D_j + \frac{\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} \le \ \alpha_i - \alpha_j) < \mathbb{P}(R_j - Q - D_j + \frac{c}{\xi} + \lambda^{-1} \le \ 0) < \frac{c}{\psi}.$$

Recall that bank i is optimizing in the payment subgame: by Lemma 9 and Lemma 4,

$$\mathbb{P}(a_j \le \alpha_i) = \mathbb{P}(L_j + \alpha_j \le \alpha_i) = \mathbb{P}(R_j - Q - D_j + \frac{\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i)}{\xi} + \lambda^{-1} \le \ \alpha_i - \alpha_j) \ge \frac{c}{\psi}.$$

This is a contradiction. Thus, $\alpha_i < N_{min} \Rightarrow \alpha_i > \alpha_j$. It further implies that when $\alpha_i < N_{min}$, $\alpha_j < N_{min}$. However, a symmetric argument shows that when $\alpha_j < N_{min}$, $\alpha_j > \alpha_i$, a contradiction. Thus, $\alpha_i \ge N_{min}$ and $\alpha_j \ge N_{min}$.

Our next goal is to show $\alpha_i = \alpha_j$. By Lemma 9 and Lemma 4,

$$\begin{aligned}\frac{c}{\psi} \leq \mathbb{P}(a_i \leq \alpha_j) =& \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq \alpha_j) \\ =& \mathbb{P}(L_i \leq \alpha_j - \alpha_i) + \mathbb{P}(N_i \leq \alpha_j)(1 - \mathbb{P}(L_i \leq \alpha_j - \alpha_i)) \\ =& \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq \alpha_j - \alpha_i - \frac{\Gamma_i(\alpha_j - \alpha_i, \alpha_i, \alpha_j)}{\xi}\right) \\ &+ \mathbb{P}(N_i \leq \alpha_j)\left(1 - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq \alpha_j - \alpha_i - \frac{\Gamma_i(\alpha_j - \alpha_i, \alpha_i, \alpha_j)}{\xi}\right)\right)\end{aligned}$$

and

$$\begin{aligned}\frac{c}{\psi} \leq \mathbb{P}(a_j \leq \alpha_i) =& \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \leq \alpha_i - \alpha_j - \frac{\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i)}{\xi}\right) \\ &+ \mathbb{P}(N_i \leq \alpha_i)\left(1 - \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \leq \alpha_i - \alpha_j - \frac{\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i)}{\xi}\right)\right).\end{aligned}$$

By Lemma 2 and Definition 1,

$$\Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i) = \int_{n \in ((\alpha_i - \alpha_j)^+, \alpha_i)} \psi \mathbb{P}(a_i \leq n - (\alpha_i - \alpha_j))\, dF_N(n) + \int_{n \in [\alpha_i, \infty)} c dF_N(n),$$

$$\Gamma_i(\alpha_j - \alpha_i, \alpha_i, \alpha_j) = \int_{n \in ((\alpha_j - \alpha_i)^+, \alpha_j)} \psi \mathbb{P}(a_j \leq n - (\alpha_j - \alpha_i))\, dF_N(n) + \int_{n \in [\alpha_j, \infty)} c dF_N(n).$$

Lemma 19. *Suppose that (R_i, D_i, N_i) and (R_j, D_j, N_j) have the same distribution. If in equilibrium bank i pays $a_i = \min((L_i + \alpha_i)^+, N_i)$ and bank j pays $a_j = \min((L_j + \alpha_j)^+, N_j)$ in the payment subgame and $\alpha_j > \alpha_i$, then*

$$\mathbb{P}(a_j \leq n) \leq \mathbb{P}(a_i \leq n).$$

for all $n \in [N_{min}, \alpha_i)$.

The proof of this lemma is very technical and is postponed in Appendix G.1. In nontechnical terms, when $\alpha_j > \alpha_i$ bank j pays more than bank i. Therefore, $\Gamma_j(y, \alpha_j, \alpha_i)$ is larger than $\Gamma_i(y, \alpha_i, \alpha_j)$ for y in some relevant range. Thus, bank j quotes higher funding rates and lends out less liquidity in the funding market, so L_j is higher than L_i. This confirms that $a_j = \min((L_j + \alpha_j)^+, N_j)$ is larger than $a_i = \min((L_i + \alpha_i)^+, N_i)$, so $\mathbb{P}(a_j \leq n) \leq \mathbb{P}(a_i \leq n)$.

With Lemma 19 we are ready to show that $\alpha_j - \alpha_i = 0$. Suppose that $\alpha_j - \alpha_i > 0$. By

Lemma 19,

$$\begin{aligned}\Gamma_i(0,\alpha_i,\alpha_j) = & \int\limits_{n\in[N_{min},\alpha_i)} \psi\mathbb{P}(a_j \le n)\, dF_N(n) + \int\limits_{n\in[\alpha_i,\infty)} c dF_N(n) \\ \le & \int\limits_{n\in[N_{min},\alpha_i)} \psi\mathbb{P}(a_i \le n)\, dF_N(n) + \int\limits_{n\in[\alpha_i,\infty)} c dF_N(n) \\ \le & \int\limits_{n\in[N_{min},\alpha_i)} \psi\mathbb{P}(a_i \le n + (\alpha_j - \alpha_i))\, dF_N(n) + \int\limits_{n\in[\alpha_i,\infty)} c dF_N(n) \\ = & \ \Gamma_j(\alpha_i - \alpha_j, \alpha_j, \alpha_i)\end{aligned}$$

Therefore,

$$\mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le -\frac{\Gamma_i(0,\alpha_i,\alpha_j)}{\xi}\right) \ge \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \le \alpha_i - \alpha_j - \frac{\Gamma_j(\alpha_i-\alpha_j,\alpha_j,\alpha_i)}{\xi}\right).$$

This implies that bank j can deviates to choose to pay $a'_j = \min((L_j + \alpha_i)^+, N_j)$ and still satisfies

$$\begin{aligned}\mathbb{P}(a_i \le \alpha_i) =& \mathbb{P}(\min((L_i+\alpha_i)^+, N_i) \le \alpha_i) \\ =& \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le -\frac{\Gamma_i(0,\alpha_i,\alpha_j)}{\xi}\right) \\ & + \mathbb{P}(N_i \le \alpha_i)\left(1 - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le -\frac{\Gamma_i(0,\alpha_i,\alpha_j)}{\xi}\right)\right) \\ \ge& \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \le \alpha_i - \alpha_j - \frac{\Gamma_j(\alpha_i-\alpha_j,\alpha_j,\alpha_i)}{\xi}\right) \\ & + \mathbb{P}(N_i \le \alpha_i)\left(1 - \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \le \alpha_i - \alpha_j - \frac{\Gamma_j(\alpha_i-\alpha_j,\alpha_j,\alpha_i)}{\xi}\right)\right) \\ & = \mathbb{P}(a_j \le \alpha_i) \ge \frac{c}{\psi}.\end{aligned}$$

In addition, $a_i < a_j \Rightarrow \mathbb{P}(a_i \le \alpha_i) < \mathbb{P}(a_i \le \alpha_j)$. This contradicts with Lemma 9.

Similarly, it cannot be the case that $\alpha_i > \alpha_j$. Thus, $\alpha_i = \alpha_j$. Let $\alpha_i = \alpha_j = \alpha$. Lemma 19 also implies that $\Gamma_i = \Gamma_j = \Gamma$ for some function Γ. To sum up, the value of α and the function Γ are jointly determined in the following way:

1. Fix the function Γ in equilibrium. Then α satisfies

$$\begin{aligned}\alpha = \inf\Bigg\{\vartheta \ge N_{min},\ & \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right) + \mathbb{P}(N_i \le \vartheta) \\ & \left(1 - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \le -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)\right) \ge \frac{c}{\psi}\Bigg\}.\end{aligned}$$

When F_N is atomless, then

$$\mathbb{P}(N_i \leq \alpha) = \frac{\frac{c}{\psi} - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)}{1 - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)}. \tag{31}$$

2. Fix the value of α. Define functional

$$F_{a_{-i}}(y, \alpha, \Gamma) = F_N(y) + (1 - F_N(y))\mathbb{P}(R_j - D_j - Q + \lambda^{-1} \leq y - \alpha - \frac{\Gamma(y-\alpha, \alpha, \alpha)}{\xi}).$$

In equilibrium, when $y + \alpha > 0$,

$$\Gamma(y, \alpha, \alpha) = \int_{n \in (y^+, (y+\alpha))} \psi F_{a_{-i}}(n - y, \alpha, \Gamma)\, dF_N(n) + \int_{n \in [(y+\alpha), \infty)} c dF_N(n). \tag{32}$$

When $y + \alpha \leq 0$,

$$\Gamma(y, \alpha, \alpha) = \psi F_{a_{-i}}(-y, \alpha, \Gamma). \tag{33}$$

The equilibrium α and Γ is the fixed point of the above system of integral equations.

Lemma 20. *There is at least one pair of α and Γ that satisfies the above system.*

(Proof see Appendix G.2.)

Lemma 20 implies that there is at least one equilibrium under the no hoarding condition.

F.8 Proof of Theorem 8

First, we explore properties for any equilibrium (α, Γ) under the given conditions. When $N_i - N_{min}$ is exponentially distributed, from Eq. (31),

$$\frac{\frac{c}{\psi} - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)}{1 - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)} = \mathbb{P}(N_i \leq \alpha) = 1 - e^{-\lambda_N(\alpha - N_{min})}.$$

Thus,

$$e^{\lambda_N(\alpha - N_{min})} = \frac{1 - \mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq -\frac{\Gamma(0,\alpha,\alpha)}{\xi}\right)}{1 - \frac{c}{\psi}} > 0 \tag{34}$$

From Eq. (32) and Eq. (33), when $L \leq -\alpha$

$$\Gamma(L,\alpha,\alpha) = \psi F_{a_{-i}}(-L,\alpha,\Gamma)$$
$$= \psi \left(e^{\lambda_N(L+N_{min})}\mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq -L - \alpha - \frac{\Gamma(-L-\alpha,\alpha,\alpha)}{\xi}\right) + 1 - e^{\lambda_N(L+N_{min})}\right); \quad (35)$$

when $-\alpha < L \leq N_{min} - \alpha$,

$$\Gamma(L,\alpha,\alpha) = \int_{n\in(L^+,(L+\alpha))} \psi F_{a_{-i}}(n-L)\, dF_N(n) + \int_{n\in[(L+\alpha),\infty)} c dF_N(n) = c;$$

when $N_{min} - \alpha < L \leq N_{min}$,

$$\Gamma(L,\alpha,\alpha) = \int_{n\in[N_{min},(L+\alpha))} \psi F_{a_{-i}}(n-L)\, dF_N(n) + \int_{n\in[(L+\alpha),\infty)} c dF_N(n) < c;$$

when $N_{min} < L$

$$\Gamma(L,\alpha,\alpha) = \int_{n\in(L,(L+\alpha))} \psi F_{a_{-i}}(n-L)\, dF_N(n) + \int_{n\in[(L+\alpha),\infty)} c dF_N(n).$$

Therefore, when $-\alpha < L \leq N_{min} - \alpha$,

$$\Gamma'(L,\alpha,\alpha) = 0.$$

Fix α and consider changing L for $\Gamma(L,\alpha,\alpha)$ and $F_{a_{-i}}(L,\alpha,\Gamma)$. To simplify notations, write $\Gamma(L,\alpha,\alpha)$ as $\Gamma(L)$ and $F_{a_{-i}}(L,\alpha,\Gamma)$ as $F_{a_j}(L)$ when there is no confusion. Our next step is to transfer the integral equation into a non-standard ODE when $L > N_{min} - \alpha$. Note by assumption $F_{a_j}(L)$ is differentiable. Let $f_{a_j}(L) = F'_{a_j}(L)$.

When $N_{min} - \alpha < L \leq N_{min}$,

$$\Gamma'(L) = \int_{n\in[N_{min},(L+\alpha))} -\psi f_{a_j}(n-L)\, dF_N(n) + \psi F_{a_j}(\alpha) f_N(L+\alpha) - c f_N(L+\alpha)$$
$$= \int_{n\in[N_{min},(L+\alpha))} -\psi f_{a_j}(n-L)\, \lambda_N e^{-\lambda_N(n-N_{min})} dn.$$

Since

$$\frac{d}{dn}\psi F_{a_j}(n-L)\lambda_N e^{-\lambda_N(n-N_{min})} = -\psi F_{a_j}(n-L)\lambda_N^2 e^{-\lambda_N(n-N_{min})} + \psi f_{a_j}(n-L)\lambda_N e^{-\lambda_N(n-N_{min})},$$

we have

$$\Gamma'(L) = -\int\limits_{n\in[N_{min},(L+\alpha))} \frac{d}{dn}\psi F_{a_j}(n-L)\lambda_N e^{-\lambda_N(n-N_{min})}dn - \int\limits_{n\in[N_{min},(L+\alpha))} \psi F_{a_j}(n-L)\lambda_N^2 e^{-\lambda_N(n-N_{min})}dn$$
$$= \psi F_{a_j}(N_{min}-L)\lambda_N - \lambda_N\Gamma(L).$$

When $0 < L \leq N_{min}$, $N_{min} - L < N_{min}$ and $\Gamma(N_{min} - L - \alpha) = c$, so

$$\begin{aligned}\Gamma'(L) =& -\lambda_N\Gamma(L) + \psi\lambda_N\mathbb{P}(R_j - D_j - Q + \lambda^{-1} \leq N_{min} - L - \alpha - \frac{\Gamma(N_{min}-L-\alpha)}{\xi})\\ =& -\lambda_N\Gamma(L) + \psi\lambda_N\mathbb{P}(R_j - D_j - Q + \lambda^{-1} \leq N_{min} - L - \alpha - \frac{c}{\xi}).\end{aligned}$$

When $N_{min} - \alpha < L \leq 0$, $N_{min} - L \geq N_{min}$, so

$$\begin{aligned}\Gamma'(L) =& -\lambda_N\Gamma(L) + \psi\lambda_N F_{a_j}(N_{min} - L)\\ =& -\lambda_N\Gamma(L) + \psi\lambda_N((1 - e^{\lambda_N L}) + e^{\lambda_N L}\mathbb{P}(R_j - D_j - Q + \lambda^{-1} \leq N_{min} - L - \alpha - \frac{\Gamma(N_{min}-L-\alpha)}{\xi})).\end{aligned}$$

Define $H(k) = e^{\lambda_N k}\Gamma(k)$. Then

$$\begin{aligned}H'(k) =& e^{\lambda_N k}\Gamma'(k) + \lambda_N e^{\lambda_N k}\Gamma(k) = e^{\lambda_N k}\psi\lambda_N F_{a_j}(N_{min} - k)\\ =& e^{\lambda_N k}\psi\lambda_N\big((1 - e^{\lambda_N k}) + e^{\lambda_N k}\mathbb{P}(R_j - D_j - Q + \lambda^{-1} \leq N_{min} - k - \alpha - \frac{\Gamma(N_{min}-k-\alpha)}{\xi})\big)\end{aligned} \tag{36}$$

so

$$H(L) = e^{\lambda_N L}\Gamma(L) = \Gamma(0) + \psi\lambda_N\int_0^L e^{\lambda_N k}F_{a_j}(N_{min} - k)dk.$$

Thus $\Gamma(L)$ is uniquely determined by $\Gamma(0)$ when $0 < L \leq N_{min}$. Next, when $N_{min} < L$,

$$
\begin{aligned}
\Gamma'(L) = & \int_{n\in[L,(L+\alpha))} -\psi f_{a_j}(n-L)\lambda_N e^{-\lambda_N(n-N_{min})}dn - \psi F_{a_j}(0) f_N(L) \\
= & -\int_{n\in[L,(L+\alpha))} \frac{d}{dn}\psi F_{a_j}(n-L)\lambda_N e^{-\lambda_N(n-N_{min})}dn \\
& -\int_{n\in[L,(L+\alpha))} \psi F_{a_j}(n-L)\lambda_N^2 e^{-\lambda_N(n-N_{min})}dn - \psi F_{a_j}(0) f_N(L) \\
= & \, \psi F_{a_j}(L-L)\lambda_N e^{-\lambda_N(L-N_{min})} - \psi F_{a_j}(\alpha)\lambda_N e^{-\lambda_N(L+\alpha-N_{min})} - \psi F_{a_j}(0) f_N(L) \\
& -\lambda_N \int_{n\in[N_{min},(L+\alpha))} \psi F_{a_j}(n-L)\lambda_N e^{-\lambda_N(n-N_{min})}dn \\
= & -c\lambda_N e^{-\lambda_N(L+\alpha-N_{min})} - \lambda_N\Gamma(L) + \lambda_N c e^{-\lambda_N(L+\alpha-N_{min})} \\
= & -\lambda_N\Gamma(L).
\end{aligned}
$$

Thus, $\Gamma(L)$ is uniquely determined by $\Gamma(N_{min})$ when $L > N_{min}$. Therefore, $\Gamma(0)$ uniquely determines $\Gamma(L)$ for all $L \geq 0$. By Eq. (35), $\Gamma(L)$ for $L \geq 0$ in turn uniquely determines the value of $\Gamma(L)$ for $L \leq -\alpha$. Thus it suffices to show that the system of integral equations has a unique solution $\Gamma(L)$ for $L \in [N_{min} - \alpha, 0]$.

Suppose there are two equilibria, characterized by (α_1, Γ_1) and (α_2, Γ_2). Assume that $\Gamma_1(0, \alpha_1, \alpha_1) > \Gamma_2(0, \alpha_2, \alpha_2)$, then it follows that $\alpha_1 > \alpha_2$. We want to establish a contradiction. Define $H_1(L) = e^{\lambda_N L}\Gamma_1(L)$ and $H_2(L) = e^{\lambda_N L}\Gamma_2(L)$. Then $H_1(0) > H_2(0)$ and

$$
\begin{aligned}
H_1(N_{min} - \alpha_1) &= e^{\lambda_N(N_{min}-\alpha_1)}c \leq e^{\lambda_N(N_{min}-\alpha_2)}c = H_2(N_{min} - \alpha_2), \\
H_1(N_{min} - \alpha_2) &= e^{\lambda_N(N_{min}-\alpha_2)}\Gamma_1(N_{min} - \alpha_2) \leq e^{\lambda_N(N_{min}-\alpha_2)}c = H_2(N_{min} - \alpha_2).
\end{aligned}
$$

Similarly, we can show $H_1'(N_{min} - \alpha_2) < H_2'(N_{min} - \alpha_2)$ and $H_1'(0) < H_2'(0)$. Note that H_1' and H_2' are continuous.

Since $H_1(N_{min} - \alpha_2) \leq H_2(N_{min} - \alpha_2)$ and $H_1(0) > H_2(0)$, there must be $x \in (N_{min} - \alpha_2, 0)$ such that $H_1'(x) > H_2'(x)$. Then there exists t_1 such that $H_1'(N_{min} - \alpha_2 + x) < H_2'(N_{min} - \alpha_2 + x), \forall 0 \leq x < t_1$ and $H_1'(N_{min} - \alpha_2 + t_1) = H_2'(N_{min} - \alpha_2 + t_1)$. Since $H_1'(0) < H_2'(0)$, there also exists t_0 such that $H_1'(x) < H_2'(x), \forall 0 \geq x > -t_0$ and $H_1'(-t_0) = H_2'(-t_0)$.

First, $-t_0 > \frac{N_{min}-\alpha_2}{2}$. If not, then $\Gamma_1(x) > \Gamma_1(x), \forall 0 \geq x > \frac{N_{min}-\alpha_2}{2}$. This implies $H_1'(x) < H_2'(x), \forall N_{min} - \alpha_2 \leq x \leq \frac{N_{min}-\alpha_2}{2}$, so $H_1(N_{min} - \alpha_2) > H_2(N_{min} - \alpha_2)$, a contradiction. Second, by $H_1'(N_{min} - \alpha_2 + t_1) = H_2'(N_{min} - \alpha_2 + t_1)$ and Eq. (36),

$$
\Gamma_1(-t_1 - (\alpha_1 - \alpha_2)) + \xi(\alpha_1 - \alpha_2) = \Gamma_2(-t_1).
$$

Also $\Gamma_1(-x-(\alpha_1-\alpha_2))+\xi(\alpha_1-\alpha_2)>\Gamma_2(-x), \forall 0\le x<t_1$. Similarly,

$$\Gamma_1(N_{min}-\alpha_2+t_0-(\alpha_1-\alpha_2))+\xi(\alpha_1-\alpha_2)=\Gamma_2(N_{min}-\alpha_2+t_0)$$

and $\Gamma_1(N_{min}-\alpha_2+x-(\alpha_1-\alpha_2))+\xi(\alpha_1-\alpha_2)<\Gamma_2(N_{min}-\alpha_2+x), \forall 0\le x<t_0$. Thus $-t_1<-t_0$.

Since $H_1(0)>H_2(0)$ and $H_1'(x)<H_2'(x), \forall 0\ge x>-t_0$, $H_1(-t_0)>H_2(-t_0)$ and $\Gamma_1(-t_0)-\Gamma_2(-t_0)>0$. Let $J_1=\sup\{\Gamma_1(x)-\Gamma_2(x)\ \big|\ x\in[-t_1,-t_2]\}$. By continuity, $\exists t_2\in[t_0,t_1]$ such that $\Gamma_1(-t_2)-\Gamma_2(-t_2)=J_1$ and $H_1'(-t_2)=H_2'(-t_2)$. Since $\Gamma_1(-t_1-(\alpha_1-\alpha_2))+\xi(\alpha_1-\alpha_2)=\Gamma_2(-t_1)$,

$$\begin{aligned}H_2(-t_1)-H_1(-t_1)=&e^{-\lambda_N t_1}(\Gamma_2(-t_1)-\Gamma_1(-t_1))\\ \ge&e^{-\lambda_N t_1}(\Gamma_2(-t_1)-\Gamma_1(-t_1-(\alpha_1-\alpha_2)))=e^{-\lambda_N t_1}\xi(\alpha_1-\alpha_2).\end{aligned}$$

Then by mean value theorem, $\exists t_3\in(t_2,t_1)$ such that

$$\begin{aligned}(t_1-t_2)(H_1'(-t_3)-H_2'(-t_3))=&(H_1(-t_2)-H_2(-t_2))-(H_1(-t_1)-H_2(-t_1))\\ \ge&e^{-\lambda_N t_1}\xi(\alpha_1-\alpha_2)+e^{-\lambda_N t_2}J_1.\end{aligned}$$

By Eq. (36),

$$\begin{aligned}H_1'(-t_3)-H_2'(-t_3)=\psi\lambda_N e^{-2\lambda_N t_3}\Big(&F_{RD}(-\lambda^{-1}+N_{min}-\alpha_1+t_3-\frac{\Gamma_1(N_{min}-\alpha_1+t_3)}{\xi})\\ &-F_{RD}(-\lambda^{-1}+N_{min}-\alpha_2+t_3-\frac{\Gamma_2(N_{min}-\alpha_2+t_3)}{\xi})\Big),\end{aligned}$$

Then by mean value theorem the previous two equations imply that

$$\begin{aligned}(-\alpha_1-\frac{\Gamma_1(N_{min}-\alpha_1+t_3)}{\xi}+\alpha_2+\frac{\Gamma_2(N_{min}-\alpha_2+t_3)}{\xi})f_{RD}(N_{min}-\lambda^{-1}-\alpha_2-\frac{\Gamma_2(N_{min}-\alpha_2+t_3)}{\xi}+\bar{s})\\ \ge\frac{e^{\lambda_N(2t_3-t_1)}\xi(\alpha_1-\alpha_2)+e^{\lambda_N(2t_3-t_2)}J_1}{\psi\lambda_N(t_1-t_2)}\end{aligned}$$

for some $\bar{s}\in[0,(-\alpha_1-\frac{\Gamma_1(N_{min}-\alpha_1+t_3)}{\xi}-(-\alpha_2-\frac{\Gamma_2(N_{min}-\alpha_2+t_3)}{\xi}))]$. Since $f_{RD}(t)\le f_{RD}^m$ when $t\in(-\infty,0]$,

$$\frac{\Gamma_2(N_{min}-\alpha_2+t_3)-\Gamma_1(N_{min}-\alpha_1+t_3)}{\xi}\ge\alpha_1-\alpha_2+\frac{e^{\lambda_N(2t_3-t_1)}\xi(\alpha_1-\alpha_2)+e^{\lambda_N(2t_3-t_2)}J_1}{f_{RD}^m\ \ \psi\lambda_N(t_1-t_2)}.$$

Again by mean value theorem, $\exists t_4\in(t_2,t_3)$ such that

$$\begin{aligned}&\Gamma_2(N_{min}-\alpha_2+t_3)-\Gamma_1(N_{min}-\alpha_1+t_3)\\ =&\Gamma_2(N_{min}-\alpha_2+t_2)-\Gamma_1(N_{min}-\alpha_1+t_2)+(t_3-t_2)(\Gamma_2'(N_{min}-\alpha_2+t_4)-\Gamma_1'(N_{min}-\alpha_1+t_4))\\ =&\xi(\alpha_1-\alpha_2)+(t_3-t_2)(\Gamma_2'(N_{min}-\alpha_2+t_4)-\Gamma_1'(N_{min}-\alpha_1+t_4)).\end{aligned}$$

Thus,

$$\psi\lambda_N\Big(e^{\lambda_N(N_{min}-\alpha_1+t_4)} - e^{\lambda_N(N_{min}-\alpha_2+t_4)} + e^{\lambda_N(N_{min}-\alpha_2+t_4)}F_{RD}(-t_4 - \frac{\Gamma_2(-t_4)}{\xi} - \lambda^{-1})$$
$$- e^{\lambda_N(N_{min}-\alpha_1+t_4)}F_{RD}(-t_4 - \frac{\Gamma_1(-t_4)}{\xi} - \lambda^{-1})\Big) \geq \frac{e^{\lambda_N(2t_3-t_1)}\xi^2(\alpha_1-\alpha_2) + \xi e^{\lambda_N(2t_3-t_2)}J_1}{f_{RD}^m\ (t_3-t_2)\psi\lambda_N(t_1-t_2)}.$$

In other words,

$$e^{\lambda_N(N_{min}-\alpha_1+t_4)}(F_{RD}(-t_4 - \frac{\Gamma_2(-t_4)}{\xi} - \lambda^{-1}) - F_{RD}(-t_4 - \frac{\Gamma_1(-t_4)}{\xi} - \lambda^{-1})) \geq$$
$$(e^{\lambda_N(N_{min}-\alpha_2+t_4)} - e^{\lambda_N(N_{min}-\alpha_1+t_4)})(1 - F_{RD}(-t_4 - \frac{\Gamma_2(-t_4)}{\xi} - \lambda^{-1}))$$
$$+ \frac{e^{\lambda_N(2t_3-t_1)}\xi^2(\alpha_1-\alpha_2) + \xi e^{\lambda_N(2t_3-t_2)}J_1}{f_{RD}^m\ (t_3-t_2)\psi^2\lambda_N^2(t_1-t_2)}$$

Thus,

$$\left(\frac{\Gamma_1(-t_4)}{\xi} - \frac{\Gamma_2(-t_4)}{\xi}\right) f_{RD}^m > \frac{e^{\lambda_N(2t_3-t_1-t_4+\alpha_1-N_{min})}\xi^2(\alpha_1-\alpha_2) + \xi e^{\lambda_N(2t_3-t_2-t_4+\alpha_1-N_{min})}J_1}{f_{RD}^m\ (t_3-t_2)\psi^2\lambda_N^2(t_1-t_2)}.$$

Note that

$$\frac{e^{\lambda_N(t_3-t_2-t_4+\alpha_1-N_{min})}}{(t_3-t_2)\lambda_N^2(t_1-t_2)} = e^{\lambda_N(t_3-t_4)}\frac{e^{\lambda_N(t_3-t_2)}}{(t_3-t_2)\lambda_N}\frac{e^{\lambda_N(\alpha_1-N_{min})}}{(t_1-t_2)\lambda_N} > 2e^2.$$

Thus,

$$\Gamma_1(-t_2) - \Gamma_2(-t_2) \geq \Gamma_1(-t_4) - \Gamma_2(-t_4) > (\frac{1}{f_{RD}^m})^2\frac{2e^2\xi^2 J_1}{\psi^2} > J_1.$$

The last line follows from the assumption of Theorem 8. Then we have $J_1 = \Gamma_1(-t_2) - \Gamma_2(-t_2) > J_1$, a contradiction. Thus, there cannot be two (α_1, Γ_1) and (α_2, Γ_2) that solves the system of integral equations defined in Theorem 7. In other words, the equilibrium is unique.

F.9 Proof of Theorem 9

For simplicity, I suppress notation and write $\Gamma(y)$ as $\Gamma(y, \alpha, \alpha)$. Repo trading is characterized by Eq. (12). Immediately, we have the following lemma.

Lemma 21. *$\mathscr{S}(D)$ is almost everywhere differentiable, and*

$$\mathscr{S}(D) = \mathscr{S}(D_{min}) + \int_{D_{min}}^{D} \mathscr{S}'(\theta)\, d\theta.$$

Proof. Consider $\theta_1 < \theta_2$ in the support of D_j. By *Lemma* 3, $\mathscr{S}(\theta_1) < \mathscr{S}(\theta_2)$ and

$$\mathscr{S}(\theta_1) = \inf\left\{s : \Gamma(R_i - Q - s) + \xi s \geq \xi(\theta_1 - \lambda^{-1})\right\}$$
$$\mathscr{S}(\theta_2) = \inf\left\{s : \Gamma(R_i - Q - s) + \xi s \geq \xi(\theta_2 - \lambda^{-1})\right\}.$$

Since $\Gamma(R_i - Q - s)$ is right continuous,

$$\Gamma(R_i - Q - \mathscr{S}(\theta_1)) + \xi\mathscr{S}(\theta_1) \geq \xi(\theta_1 - \lambda^{-1})$$
$$\Gamma(R_i - Q - \mathscr{S}(\theta_2)) + \xi\mathscr{S}(\theta_2) \geq \xi(\theta_2 - \lambda^{-1}).$$

We notice that

$$\Gamma(R_i - Q - (\mathscr{S}(\theta_1) + \theta_2 - \theta_1)) + \xi(\mathscr{S}(\theta_1) + \theta_2 - \theta_1) \geq \xi(\theta_1 + \theta_2 - \theta_1 - \lambda^{-1}) = \xi(\theta_2 - \lambda^{-1}).$$

Since $\mathscr{S}(\theta_2) = \inf\left\{s : \Gamma(R_i - Q - s) + \xi s \geq \xi(\theta_2 - \lambda^{-1})\right\}$,

$$\mathscr{S}(\theta_1) \leq \mathscr{S}(\theta_2) \quad \leq \mathscr{S}(\theta_1) + \theta_2 - \theta_1.$$

Thus,

$$|\mathscr{S}(\theta_2) - \mathscr{S}(\theta_1)| \leq \theta_2 - \theta_1.$$

Hence $\mathscr{S}(\theta)$ is absolute continuous. By Lebesgue differentiation theorem, $\mathscr{S}(D)$ is differentiable almost everywhere, and

$$\mathscr{S}(D) = \mathscr{S}(D_{min}) + \int_{D_{min}}^{D} \mathscr{S}'(\theta)\, d\theta.$$

□

By Eq. (12),

$$\Gamma(R_i - Q - \mathscr{S}(D)) \geq \xi(D - \mathscr{S}(D) - \lambda^{-1})$$
$$\mathcal{T}(D) = -\frac{\xi}{2}\Big(D - \mathscr{S}(D)\Big)^2 + \frac{\xi}{2}{D_{min}}^2 + \int_{D_{min}}^{D} \xi(x - \mathscr{S}(x))\, dx$$
$$r_i = \frac{\mathcal{T}(D)}{\mathscr{S}(D)}.$$

By Lemma 21, $\mathscr{S}(D)$ is differentiable for all $D \in [D_{min}, \infty)$, so $\mathcal{T}(D)$ is differentiable almost everywhere (a.e):

$$\mathcal{T}'(D) = \xi(D - \mathscr{S})\mathscr{S}'(D) \qquad \text{a.e.}$$

Lemma 22. *Suppose that for some θ_1, $\Gamma(R_i - Q - \mathscr{S}(\theta_1)) > \xi(\theta_1 - \mathscr{S}(\theta_1) - \lambda^{-1})$, then there is an $\theta_2 > \theta_1$ such that*

$$\begin{aligned}\Gamma(R_i - Q - \mathscr{S}(\theta_2)) =& \xi(\theta_2 - \mathscr{S}(\theta_2) - \lambda^{-1})\\ \mathscr{S}(\theta_2) =& \mathscr{S}(\theta_1)\\ \mathscr{S}'(x) =& 0, \forall x \in (\theta_1, \theta_2).\end{aligned}$$

Moreover, there are at most countably many θ such that $\Gamma(R_i - Q - \mathscr{S}(\theta_1)) > \xi(\theta_1 - \mathscr{S}(\theta_1) - \lambda^{-1})$.

Proof. Suppose that for some θ_1, $\Gamma(R_i - Q - \mathscr{S}(\theta_1)) > \xi(\theta_1 - \mathscr{S}(\theta_1) - \lambda^{-1})$, then

$$\Gamma(R_i - Q - \mathscr{S}(\theta_1)) + \xi\mathscr{S}(\theta_1) > \xi(\theta_1 - \lambda^{-1}).$$

Since $\xi(x - \lambda^{-1})$ is strictly increasing in x, there is a $\theta_2 > \theta_1$ such that

$$\Gamma(R_i - Q - \mathscr{S}(\theta_1)) + \xi\mathscr{S}(\theta_1) = \xi(\theta_1 - \lambda^{-1}).$$

For any $x \in [\theta_1, \theta_2]$, $\Gamma(R_i - Q - \mathscr{S}(\theta_1)) + \xi\mathscr{S}(\theta_1) \geq \xi(x - \lambda^{-1})$ and for any $s < \mathscr{S}(\theta_1)$ $\Gamma(R_i - Q - s) + \xi s < \xi(\theta_1 - \lambda^{-1}) < \xi(x - \lambda^{-1})$. Thus, $\mathscr{S}(x) = \mathscr{S}(\theta_1), \quad \forall x \in [\theta_1, \theta_2]$. It follows that $\mathscr{S}'(x) = 0, \forall x \in (\theta_1, \theta_2)$. Finally, the function $x \mapsto \Gamma(R_i - Q - \mathscr{S}(x)) + \xi\mathscr{S}(x)$ is strictly increasing in x, so it has at most countably many discontinuity. □

Following Lemma 22, $\xi(x - \mathscr{S}(x))\mathscr{S}'(x) = \Big(\Gamma(R_i - Q - \mathscr{S}(x)) + \xi\lambda^{-1}\Big)\mathscr{S}'(x)$ almost everywhere. Thus,

$$\begin{aligned}\mathcal{T}(D) =& \mathcal{T}(D_{min}) + \int_{D_{min}}^{D} \xi(x - \mathscr{S}(x))\mathscr{S}'(x)\, dx\\ =& \mathcal{T}(D_{min}) + \int_{D_{min}}^{D} \Big(V'_-(R_i - Q - \mathscr{S}(x)) + \xi\lambda^{-1}\Big)\mathscr{S}'(x)\, dx\\ =& \mathcal{T}(D_{min}) + \int_{\mathscr{S}(D_{min})}^{\mathscr{S}(D)} \Gamma(R_i - Q - s)\, ds + \xi\,\lambda^{-1}\,(\mathscr{S}(D) - \mathscr{S}(D_{min})).\end{aligned} \tag{37}$$

Plug the calculation of $\mathcal{T}(D)$ into Eq. (12) and we get the desired result.

F.10 Proof of Theorem 10

By Lemma 3, Theorem 6 and Theorem 7, the equilibrium outcome variables S_i, T_i and r_i of the trading game depends on the realizations of R_i, D_i and m. Let $\mathcal{S}(\zeta, \mathscr{D}(\mathcal{S}^*, m), m)$ and $\mathscr{T}(\zeta, \mathscr{D}(\mathcal{S}^*, m), m)$ be the equilibrium amount of financing and total transfer respectively, in the trading game in the state of the world such that the realization of R_i is ζ and the realization of D_i is $\mathscr{D}(\mathcal{S}^*, m)$. Obviously, $\mathscr{D}(\mathcal{S}^*, m)$ has to satisfy $\mathcal{S}(\zeta, \mathscr{D}(\mathcal{S}^*, m), m) = \mathcal{S}^*$.

Abuse notations and write $\Gamma(y,\alpha,\alpha)$ as $\Gamma(y\ ;\ m)$ to denote the marginal value of liquidity at y when the liquidity stress index is m. Let $\ell = \zeta - Q - \mathcal{S}^*$. By Eq. (37), under the macroeconomic conditions indexed by m,

$$\begin{aligned}\mathscr{T}(\zeta,\mathscr{D}(\mathcal{S}^*,m),m) =&\mathscr{T}(\zeta,D_{min},m) + \int\limits_{\mathcal{S}(\zeta,D_{min},m)}^{\mathcal{S}(\zeta,\mathscr{D}(\mathcal{S}^*,m),m)} \Gamma(\zeta - Q - s, m)\, ds \\ &+ \frac{\xi^m}{\lambda^m}\big(\mathcal{S}(\zeta,\mathscr{D}(\mathcal{S}^*,m),m) - \mathcal{S}(\zeta,D_{min},m)\big).\end{aligned} \tag{38}$$

By Lemma 3,

$$\begin{aligned}\mathscr{T}(\zeta,D_{min},m) =& -\frac{\xi^m}{2}\Big(D_{min} - \mathcal{S}(\zeta,D_{min},m)\Big)^2 + \frac{\xi^m}{2}{D_{min}}^2 \\ =&\,\xi^m D_{min}\mathcal{S}(\zeta,D_{min},m) - \frac{\xi^m}{2}\mathcal{S}(\zeta,D_{min},m)^2.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathscr{T}(\zeta,D_{min},m) - \frac{\xi^m}{\lambda^m}\mathcal{S}(\zeta,D_{min},m) \\ &\quad= \ (\xi^m D_{min} - \frac{\xi^m}{\lambda^m})\mathcal{S}(\zeta,D_{min},m) - \frac{\xi^m}{2}\mathcal{S}(\zeta,D_{min},m)^2.\end{aligned} \tag{39}$$

By definition,

$$r^*(\zeta,\mathcal{S}^*,m) = \frac{\mathscr{T}(\zeta,\mathscr{D}(\mathcal{S}^*,m),m)}{\mathcal{S}^*}.$$

Combining Eqs. (38) and (39), we can simplify the equilibrium funding rate as

$$\begin{aligned}r^*(\zeta,\mathcal{S}^*,m) =&\frac{1}{\mathcal{S}^*}\Bigg((\xi^m D_{min} - \frac{\xi^m}{\lambda^m})\mathcal{S}(\zeta,D_{min},m) - \frac{\xi^m}{2}\mathcal{S}(\zeta,D_{min},m)^2 \\ &+ \int\limits_{\mathcal{S}(\zeta,D_{min},m)}^{\zeta - Q^m - \ell} \Gamma(\zeta - Q^m - s\ ;\ m)\, ds\Bigg) + \frac{\xi^m}{\lambda^m}.\end{aligned}$$

To simplify notations, let $\{F_R(\cdot),F_N(\cdot),Q,\lambda,D_{min},c,\psi,\xi\} = \{F_R^0(\cdot),F_N^0(\cdot),Q^0,\lambda^0,D_{min}^0,c^0,\psi^0,\xi^0\}$. When $m=0$, there could be multiple equilibria. For the following analysis, we select the equilibrium with the max $\alpha = N_{min}$, because, by continuity, this equilibrium is the limiting equilibrium of a sequence of economies under no hoarding condition such that $m \uparrow 0$. In other words, pick the equilibrium such that $r^*(\zeta,\mathcal{S}^*,0) = \lim_{\epsilon^m \downarrow 0} r^*(\zeta,\mathcal{S}^*,-\epsilon^m)$.

Fix some $\epsilon^m > 0$. Since all the macroeconomic conditions considered here are close with respect to liquidity stress index, there exists some constants $\mathcal{O},\overline{\epsilon^m} > 0$ such that for all $\epsilon^m < \overline{\epsilon^m}$,

$$
\begin{aligned}
&\Big(r^*(\zeta,\mathcal{S}^*,\epsilon^m) - r^*(\zeta,\mathcal{S}^*,0)\Big)\mathcal{S}^* \\
\geq\ & (\xi D_{min} - \xi\lambda^{-1})\mathcal{S}(\zeta, D_{min}, \epsilon^m) - \frac{\xi}{2}\mathcal{S}(\zeta, D_{min}, \epsilon^m)^2 - (\xi D_{min} - \xi\lambda^{-1})\mathcal{S}(\zeta, D_{min}, 0) + \frac{\xi}{2}\mathcal{S}(\zeta, D_{min}, 0)^2 \\
&+ \int_{\mathcal{S}(\zeta,D_{min}\,;\,0)}^{\mathcal{S}^*} \Gamma(\zeta - Q - s\,;\,\epsilon^m) - \Gamma(\zeta - Q - s\,;\,0)\, ds + \int_{\mathcal{S}(\zeta,D_{min}\,;\,\epsilon)}^{\mathcal{S}(\zeta,D_{min}\,;\,0)} \Gamma(\zeta - Q - s\,;\,\epsilon^m)\, ds - \mathcal{O}\epsilon^m \\
=\ & \int_{\mathcal{S}(\zeta,D_{min}\,;\,0)}^{\mathcal{S}^*} \Gamma(\zeta - Q - s\,;\,\epsilon^m) - \Gamma(\zeta - Q - s\,;\,0)\, ds - \mathcal{O}\epsilon^m \\
&+ \int_{\mathcal{S}(\zeta,D_{min}\,;\,\epsilon)}^{\mathcal{S}(\zeta,D_{min}\,;\,0)} -(\xi D_{min} - \xi\lambda^{-1}) + \xi s + \Gamma(\zeta - Q - s\,;\,\epsilon^m)\, ds.
\end{aligned}
$$

Since

$$
\Gamma\Big(\zeta - Q - \mathcal{S}(\zeta, D_{min}\,;\,\epsilon^m)\,;\,\epsilon^m\Big) \geq \xi(D_{min} - \lambda^{-1}) - \xi\mathcal{S}(\zeta, D_{min}\,;\,\epsilon^m)
$$

and the mapping $s \mapsto \Gamma(\zeta - Q - s\,;\,\gamma) + \xi s$ is monotonically increasing in s for any γ,

$$
\int_{\mathcal{S}(\zeta,D_{min}\,;\,\epsilon)}^{\mathcal{S}(\zeta,D_{min}\,;\,0)} -(\xi D_{min} - \xi\lambda^{-1}) + \xi s + \Gamma(\zeta - Q - s\,;\,\epsilon)\, ds \geq 0.
$$

Thus,

$$
\Big(r^*(\zeta,\mathcal{S}^*,\epsilon) - r^*(\zeta,\mathcal{S}^*,0)\Big)\mathcal{S}^* \geq \int_{\mathcal{S}(\zeta,D_{min}\,;\,0)}^{\mathcal{S}^*} \Gamma(\zeta - Q - s\,;\,\epsilon) - \Gamma(\zeta - Q - s\,;\,0)\, ds - \mathcal{O}\epsilon^m. \tag{40}
$$

Lemma 23. *For all $\epsilon^m < \overline{\epsilon^m}$, there is some $\delta_0(\zeta, s) \geq 0$ and constant $\mathcal{O}^1 \geq 0$ such that*

$$
\Gamma(\zeta - Q - s\,;\,\epsilon^m) - \Gamma(\zeta - Q - s\,;\,0) > \delta_0(\zeta, s) - \mathcal{O}^1\epsilon^m.
$$

In addition, whenever $\zeta - Q - s \neq 0$, $\delta_0(\zeta, s) > 0$.

Lemma 23 and Eq. (40) imply that for all $\epsilon^m < \overline{\epsilon^m}$, there is some $\delta(\zeta, \mathcal{S}^*)$ such that $r^*(\zeta, \mathcal{S}^*, \epsilon^m) - r^*(\zeta, \mathcal{S}^*, 0) > \delta(\zeta, \mathcal{S}^*) - \mathcal{O}^2\epsilon^m$ for some constant $\mathcal{O}^2$. This proves the Theorem 10.

Proof of Lemma 23. Recall that under no hoarding condition, $a_j^* = \min((L_j + \alpha)^+, N_j)$ for some $\alpha \geq N_{min}$. Under no hoarding condition, $a_j^* = \min(L_j^+, N_j)$. We discuss $\Gamma(y\,;\,\epsilon) -$

$\Gamma(y\,;\,0)$ case by case.

First, when $y \geq 0$, by directly calculation the assumption that two macroeconomic conditions are close with respect to liquidity stress index, there is some constant $\mathcal{O}^3 > 0$ such that

$$\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) \geq \int_{y}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n-y)\, d\, F_N(n) - \mathcal{O}^3\epsilon^m.$$

When $y \geq N_{min}$,

$$\begin{aligned}
&\int_{y}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n-y)\, d\, F_N(n) \\
= &\int_{y^+}^{y+\frac{N_{min}}{2}} c - \psi\mathbb{P}(\min\{L_j+\alpha, N_j\} \leq n-y)\, d\, F_N(n) + \int_{y+\frac{N_{min}}{2}}^{y+\alpha} c - \psi\mathbb{P}(\min\{L_j+\alpha, N_j\} \leq n-y)\, d\, F_N(n) \\
\geq &\int_{y^+}^{y+\frac{N_{min}}{2}} c - \psi\mathbb{P}(L_j+\alpha \leq \frac{N_{min}}{2})\, d\, F_N(n) + \int_{y+\frac{N_{min}}{2}}^{y+\alpha} c - \psi\mathbb{P}(\min\{L_j+\alpha, N_j\} \leq n-y)\, d\, F_N(n).
\end{aligned}$$

Since $c - \psi\mathbb{P}(\min\{L_j+\alpha, N_j\} \leq \alpha) = 0$ and $\mathbb{P}(\min\{L_j+\alpha, N_j\} \leq \vartheta)$ strictly increases in ϑ, there is some $\delta_1(\zeta, y)$ such that

$$\int_{y^+}^{y+\frac{N_{min}}{2}} c - \psi\mathbb{P}(L_j+\alpha \leq \frac{N_{min}}{2})\, d\, F_N(n) > \delta_1(\zeta, y) > 0,$$

and

$$\int_{y+\frac{N_{min}}{2}}^{y+\alpha} c - \psi\mathbb{P}(\min\{L_j+\alpha, N_j\} \leq n-y)\, d\, F_N(n) \geq 0.$$

Thus,

$$\int_{y}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n-y)\, d\, F_N(n) > \delta_1(\zeta, y) > 0,$$

Hence, when $y \geq N_{min}$,

$$\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) > \delta_1(\zeta, y) - \mathcal{O}^3\epsilon^m.$$

When $N_{min} > y > 0$,

$$\begin{aligned}\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) &\geq \int_{N_{min}}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n - y)\, d\, F_N(n) - \mathcal{O}^3\epsilon^m \\ &> \int_{N_{min}}^{y+\alpha} c - \psi\mathbb{P}(\min\{(L_j+\alpha)^+, N_j\} \leq n - N_{min})\, d\, F_N(n) - \mathcal{O}^3\epsilon^m.\end{aligned}$$

When $y = 0$, clearly $\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) \geq -\mathcal{O}^3\epsilon^m$. When $0 > y \geq -\alpha$, $\exists \mathcal{O}^4 > 0$ such that

$$\begin{aligned}&\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) \\ &\geq \psi\,\mathbb{P}^{\epsilon^m}(\min(L_j^+, N_j) \leq -y) - \psi \int_0^{y+\alpha} \mathbb{P}(a_j^* \leq n - y) d\, F_N(n) - c\,\mathbb{P}(N_i > y + \alpha) - \mathcal{O}^4\epsilon^m \\ &> \delta_2(\zeta, y) + c - \psi \int_0^{y+\alpha} \mathbb{P}(a_j^* \leq n - y\,;\, T + \epsilon) d\, F_N(n) - c\,\mathbb{P}(N_i > y + \alpha) - \mathcal{O}^4\epsilon^m > \delta_2(\zeta, y) - \mathcal{O}^4\epsilon^m.\end{aligned}$$

where the second inequality derives from the fact that under liquidity hoarding condition, $\psi\,\mathbb{P}^{\epsilon^m}(L_j \leq 0) > c$. Finally, when $y < -\alpha$, then

$$\begin{aligned}\Gamma(y\,;\,\epsilon^m) - \Gamma(y\,;\,0) = \;&\psi\,\mathbb{P}^{\epsilon^m}(\min(L_j^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min((L_j+\alpha)^+, N_j) \leq -y) \\ > \;&\psi\,\mathbb{P}^{\epsilon^m}(\min(L_j^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min(L_j^+, N_j) \leq -y) \\ &+ \psi\left(\mathbb{P}(\min((L_j+\alpha)^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min(L_j^+, N_j) \leq -y)\right) \\ > \;&\psi\left(\mathbb{P}(\min((L_j+\alpha)^+, N_j) \leq -y) - \psi\,\mathbb{P}(\min((L_j)^+, N_j) \leq -y)\right) > \delta_3(\zeta, y)\end{aligned}$$

for some $\delta_3(\zeta, y) > 0$ regardless of ϵ. Let $\delta_0(\zeta, \zeta - Q - y)$ be the corresponding $\delta_i(\zeta, y)$ in each case corresponding to different y. This finishes the proof. □

F.11 Proof of Lemma 5

Similar analysis as in the proof for Lemma 1 gives the following result:

Lemma 24. *Given all other banks' strategy a_{-i}, the best response actions of bank i are of the form*

$$a_i = \min((L_i + z_i)^+, N_i),$$

where z_i is some non negative random variable with support $\mathcal{I}_z$. For any $\upsilon \in \mathcal{I}_z$,

$$\mathbb{P}(a_{-i} \leq \upsilon) \geq \frac{c}{\psi}.$$

Moreover, if there does not exist $\upsilon^\star$ such that $\mathbb{P}(a_{-i} \leq \upsilon^\star) = c/\psi$, then $\mathcal{I}_z = \{\upsilon^*\}$ is a singleton and $\upsilon^* = \inf\{\vartheta \geq 0, \mathbb{P}(a_{-i} \leq \vartheta) > \frac{c}{\psi}\}$. If there is an $\upsilon^\star$ such that $\mathbb{P}(a_{-i} \leq \upsilon^\star) = c/\psi$, then for any $\upsilon \in \mathcal{I}_z$, $\mathbb{P}(a_{-i} \leq \upsilon) = c/\psi$.

Lemma 25. *Suppose that $\prod_{j\neq i} \mathbb{P}(L_j \leq 0) > \frac{c}{\psi}$ for all i, then there is a unique pure strategy equilibrium. In this equilibrium, each bank i chooses the payment $a_i^* = \min(L_i^+, N_i)$.*

Proof. When $\prod_{j\neq i} \mathbb{P}(L_j \leq 0) > \frac{c}{\psi}$ for all i, then $\mathbb{P}(\sum_{j\neq i} \min(L_j^+, N_j) \leq 0) > \frac{c}{\psi}$. This implies that $\alpha_i = 0$. If not, assume bank i makes payment $\min((L_i + \alpha_i)^+, N_i)$ for some $\alpha_i \geq 0$. WOLG assume $\alpha_1 = \max\{\alpha_i\} > 0$. Then $\mathbb{P}(\sum_{j\neq i} \min((L_j + \alpha_j)^+, N_j) \leq (n-1)\alpha_1) \geq \mathbb{P}(\sum_{j\neq i} \min((L_j + \alpha_j)^+ - \alpha_1, N_j - \alpha_1) \leq 0) > \mathbb{P}(\sum_{j\neq i} \min(L_j^+, N_j) \leq 0) > \frac{c}{\psi}$, a contradiction. □

Lemma 26. *Suppose that $\prod_{j\neq i} \mathbb{P}(L_j \leq 0) < \frac{c}{\psi}$ for all i, then there is a unique pure strategy equilibrium. In this equilibrium, each bank i chooses the payment $a_i^* = \min((L_i + \alpha_i)^+, N_i)$. For each i, $\alpha_i > 0$ and $\mathbb{P}(\sum_{j\neq i} \min((L_j + \alpha_j)^+, N_j) \leq \alpha_i) = \frac{c}{\psi}$.*

Proof. Consider a map $\mathcal{T}^\alpha : [0, F_N^{-1}(\frac{c}{\psi})]^n \to [0, F_N^{-1}(\frac{c}{\psi})]^n$, where $\mathcal{T}^\alpha(x_1, x_2, \cdots, x_n) = (\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_n)$ such that

$$\mathfrak{a}_i = \inf \left\{ \vartheta \geq 0 : \quad \mathbb{P}(\sum_{j\neq i} \min((L_j + x_j)^+, N_j) \leq (n-1)\vartheta) \geq \frac{c}{\psi} \right\}.$$

It is easy to check that $\mathcal{T}^\alpha$ is continuous. By Schauder fixed-point theorem there is at least one fixed point $(\alpha_i)_{i=1}^n$ for $\mathcal{T}^\alpha$. When $\mathbb{P}(L_j < 0) = 0$ for all j, then $(n-1)\alpha_i \geq N_{min}$. To see that, suppose $\alpha_k = \min\{\alpha_i\} < \frac{N_{min}}{n-1}$. Then

$$\mathbb{P}(\sum_{j\neq i} \min((L_j + \alpha_j)^+, N_j) \leq (n-1)\alpha_k) = \mathbb{P}(\sum_{j\neq i} L_j + \alpha_j \leq (n-1)\alpha_k) < \mathbb{P}(\sum_{j\neq i} L_j \leq 0) < \frac{c}{\psi},$$

a contradiction. Suppose there are two different fixed points, $(\alpha_i^1)_{i=1}^n$ and $(\alpha_i^2)_{i=1}^n$, for $\mathcal{T}^\alpha$. Without loss of generality (WLOG), assume that $\alpha_1^1 < \alpha_1^2$. By assumption N_j has continuous cdf, so

$$\mathbb{P}(\min((L_2 + \alpha_2^1)^+, N_2) + \sum_{j>2} \min((L_j + \alpha_j^1)^+, N_j) \leq (n-1)\alpha_1^1) = \frac{c}{\psi}$$
$$\mathbb{P}(\min((L_2 + \alpha_2^2)^+, N_2) + \sum_{j>2} \min((L_j + \alpha_j^2)^+, N_j) \leq (n-1)\alpha_1^2) = \frac{c}{\psi}.$$

Thus, it cannot be the case that $\alpha_j^2 = \alpha_j^1 = 0$ for all $j > 1$. Assume that $\alpha_2^2, \alpha_2^1 > 0$, then

$$\mathbb{P}(\min((L_1 + \alpha_1^1)^+, N_1) + \sum_{j>2} \min((L_j + \alpha_j^1)^+, N_j) \leq (n-1)\alpha_2^1) = \frac{c}{\psi}$$
$$\mathbb{P}(\min((L_1 + \alpha_1^2)^+, N_1) + \sum_{j>2} \min((L_j + \alpha_j^2)^+, N_j) \leq (n-1)\alpha_2^2) = \frac{c}{\psi}.$$

Since $\{L_j\}$, $\{N_j\}$ are i.i.d, the above equations imply that $\alpha_2^2 \geq \alpha_2^1$. Replacing α_2^2, α_2^1 with α_j^2, α_j^1, the same argument shows that $\alpha_j^2 \geq \alpha_j^1$. WLOG, assume that $\alpha_1^2 - \alpha_1^1 = \max_j\{\alpha_j^2 - \alpha_j^1\}$. Then

$$\mathbb{P}(\sum_{j>1} \min((L_j + \alpha_j^1)^+, N_j) \leq (n-1)\alpha_1^1) = \frac{c}{\psi} = \mathbb{P}(\sum_{j>1} \min((L_j + \alpha_j^2)^+, N_j) \leq (n-1)\alpha_1^2) \Rightarrow$$
$$\mathbb{P}(\sum_{j>1} (\min((L_j + \alpha_j^1)^+, N_j) + \alpha_1^2 - \alpha_1^1) \leq (n-1)\alpha_1^2) = \mathbb{P}(\sum_{j>1} \min((L_j + \alpha_j^2)^+, N_j) \leq (n-1)\alpha_1^2).$$

Note that $\min((L_j + \alpha_j^1)^+, N_j) + \alpha_1^2 - \alpha_1^1 \geq \min((L_j + \alpha_j^2)^+, N_j)$ almost surely. Suppose that $\mathbb{P}(L_j < 0) > 0$ for some j, then $\mathbb{P}(\min((L_j + \alpha_j^1)^+, N_j) + \alpha_1^2 - \alpha_1^1 > \min((L_j + \alpha_j^2)^+, N_j))$ and LHS < RHS. This is a contradiction. When $\mathbb{P}(L_j < 0) = 0$ for all j, then by the previous argument, $(n-1)\alpha_1^1 \geq N_{min}$ and $(n-1)\alpha_1^2 \geq N_{min}$. However, when this happens, since F_N strictly increases, LHS < RHS. This is a contradiction. □

Suppose that there is a mixed strategy equilibrium such that $a_i = \min((L_i + z_i)^+, N_i)$. Let $\mathcal{I}_z^i$ denote the support of z_i. Since $\mathcal{I}_z^i$ and $\mathcal{I}_z^j$ are bounded, let $\underline{v}_i \stackrel{def}{=} \inf\ \mathcal{I}_z^i$ and $\overline{v}_i \stackrel{def}{=} \sup\ \mathcal{I}_z^i$. At least one of $\mathcal{I}_z^i$ must have more than one element, for otherwise it is a pure strategy equilibrium. Say $\mathcal{I}_z^1$ has at least two elements, then $\overline{v}_1 > \underline{v}_1$ and for any $v \in [\underline{v}_1, \overline{v}_1), \mathbb{P}(a_{-1} \leq v) = \frac{c}{\psi}$.

Lemma 27. *There is no mixed strategy equilibrium.*

Proof. Suppose that there is a mixed strategy equilibrium. If $\underline{v}_1 \geq N_{min}$ Pick any $N_{min} \leq v_1' < v_1''$ in $\mathcal{I}_z^1$. Since F_N strictly increases, for any $j > 1$, $\mathbb{P}(N_j \leq v_1'') > \mathbb{P}(N_j \leq v_1')$. Thus, if

$$\mathbb{P}(a_{-i} \leq v_i'') = \mathbb{P}(a_{-1} \leq v_1') \Rightarrow$$
$$\mathbb{P}(\sum_{j>2} \min((L_j + z_j)^+, N_j) \leq (n-1)v_1'') = \mathbb{P}(\sum_{j>2} \min((L_j + z_j)^+, N_j) \leq (n-1)v_1').$$

This is impossible, since bank j cannot choose her strategy based on the the state variable of bank $k \neq j$. In other words, $\{\min((L_j + z_j)^+, N_j)\}$ are independent of each other. Since N_j is strictly increasing, the cdf of $\sum_{j>2} \min((L_j + z_j)^+, N_j)$ is strictly increasing. Thus, LHS > RHS. □

Finally, similar analysis as in the proof for Lemma 1 shows that when L_i are i.i.d, then banks' payment strategies are symmetric: $\alpha_i = \alpha$ for all i.

F.12 Proof of Theorem 11

Suppose that liquidity hoarding condition for n banks holds. We prove by contradiction that banks hoard liquidity in equilibrium at time 1. By Lemma 27 there is no mixed strategy equilibrium for the payment subgame. By Lemma 24 and Lemma 6, $\Gamma_i^+(y, \alpha_i, \alpha_{-i}) \leq c$ for any $y \geq 0$ when $\alpha_i = 0$; $\Gamma_i^+(y, \alpha_i, \alpha_{-i}) < c$ for any $y \geq 0$ when $\alpha_i > 0$. Assume that bank j does not hoard liquidity, i.e. bank j pays $\min(L_j^+ + \alpha_j, N_j)$ at time 1 for some $\alpha_j > 0$. WOLG, let $\alpha_j = \max\{\alpha_i\}$. Since bank j is best responding, it must still hold that

$$\mathbb{P}(a_{-j} \leq \alpha_j) = \mathbb{P}(\sum_{i \neq j} \min((L_i + \alpha_i)^+, N_i) \leq (n-1)\alpha_j) = \frac{c}{\psi} \Rightarrow$$

$$\mathbb{P}(\sum_{i \neq j} \min((L_i + \alpha_i)^+ - \alpha_j, N_i - \alpha_j) \leq 0) = \frac{c}{\psi} \Rightarrow$$

$$\prod_{i \neq j} \mathbb{P}(L_i \leq 0) = \mathbb{P}(\sum_{i \neq j} \min(L_i^+, N_i) \leq 0) < \mathbb{P}(\sum_{i \neq j} \min((L_i + \alpha_i)^+ - \alpha_j, N_i - \alpha_j) \leq 0) = \frac{c}{\psi}.$$

By Lemma 7,

$$\mathbb{P}(L_i \leq 0) = \mathbb{P}\left(R_i - D_i - Q \leq -\frac{\Gamma_i^+(0, \alpha_i, \alpha_{-i})}{\xi}\right) \geq \mathbb{P}\left(R_i - D_i - Q \leq -\frac{c}{\xi}\right) > (\frac{c}{\psi})^{1/(n-1)} \Rightarrow$$
$$\prod_{i \neq j} \mathbb{P}(L_i \leq 0) > \frac{c}{\psi},$$

a contradiction. Thus, $\alpha_i = 0$ and the marginal value of liquidity functions for all banks are the same.

F.13 Proof of Theorem 12

Suppose that Eq. (18) holds. By Lemma 27 there is no mixed strategy equilibrium for the payment subgame. Suppose that bank i pays $a_i = \min((L_i + \alpha_i)^+, N_i)$, our goal is to show $\alpha_i = \alpha_j, \forall i, j$. First, since $\frac{c}{\psi} > \prod_{i \neq j} \mathbb{P}(L_i \leq 0) = \mathbb{P}(\sum_{i \neq j} \min(L_i^+, N_i) \leq 0) \geq \mathbb{P}(\sum_{i \neq j} \min((L_i + \alpha_i)^+, N_i) \leq 0)$, $\alpha_i > 0$ for all i. Thus, by Lemma 24 and Lemma 7,

$$\frac{c}{\psi} = \mathbb{P}(a_{-j} \leq \alpha_j) = \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) + \sum_{k \neq i,j} \min((L_k + \alpha_k)^+, N_k) \leq (n-1)\alpha_j).$$

Let $\tilde{K} = \sum_{k \neq i,j} \min((L_k + \alpha_k)^+, N_k)$, then

$$\frac{c}{\psi} = \mathbb{P}(a_{-j} \leq \alpha_j) = \int_0^{(n-1)\alpha_j} \mathbb{P}(\min((L_i + \alpha_i)^+, N_i) \leq (n-1)\alpha_j - \kappa)\ dF_{\tilde{K}}(\kappa)$$

$$= \int_0^{(n-1)\alpha_j} \mathbb{P}\left(R_i - D_i - Q \leq (n-1)\alpha_j - \alpha_i - \kappa - \frac{\Gamma_i((n-1)\alpha_j - \alpha_i - \kappa, \alpha_i, \alpha_{-i})}{\xi}\right) + \mathbb{P}(N_i \leq (n-1)\alpha_j - \kappa)$$
$$\left(1 - \mathbb{P}\left(R_i - D_i - Q \leq (n-1)\alpha_j - \alpha_i - \kappa - \frac{\Gamma_i((n-1)\alpha_j - \alpha_i - \kappa, \alpha_i, \alpha_{-i})}{\xi}\right)\right)\ dF_{\tilde{K}}(\kappa)$$

and

$$\frac{c}{\psi} = \int_0^{(n-1)\alpha_i} \mathbb{P}\left(R_j - D_j - Q \leq (n-1)\alpha_i - \alpha_j - \kappa - \frac{\Gamma_j((n-1)\alpha_i - \alpha_j - \kappa, \alpha_j, \alpha_{-j})}{\xi}\right) + \mathbb{P}(N_j \leq (n-1)\alpha_i - \kappa)$$
$$\left(1 - \mathbb{P}\left(R_j - D_j - Q \leq (n-1)\alpha_i - \alpha_j - \kappa - \frac{\Gamma_j((n-1)\alpha_i - \alpha_j - \kappa, \alpha_j, \alpha_{-j})}{\xi}\right)\right)\ dF_{\tilde{K}}(\kappa)$$

Lemma 28. *Suppose that (R_i, D_i, N_i) and (R_j, D_j, N_j) have the same distribution. If in equilibrium bank i pays $a_i = \min((L_i + \alpha_i)^+, N_i)$ and bank j pays $a_j = \min((L_j + \alpha_j)^+, N_j)$ in the payment subgame and $\alpha_j > \alpha_i$, then $\mathbb{P}(a_{-i} \leq \eta) \leq \mathbb{P}(a_{-j} \leq \eta)$ for all $\eta \in [N_{min}, \alpha_j)$.*

The proof is omitted since it is similar to the proof for Lemma 19 with slight modifications. Lemma 28 implies that $\Gamma_j((n-1)\alpha_i - \alpha_j - \kappa, \alpha_j, \alpha_{-j}) \geq \Gamma_i((n-1)\alpha_j - \alpha_i - \kappa, \alpha_i, \alpha_{-i})$, for any $\kappa \in [0, (n-1)\alpha_i]$. However, this means

$$\frac{c}{\psi} > \int_0^{(n-1)\alpha_i} \mathbb{P}\left(R_i - D_i - Q \leq (n-1)\alpha_j - \alpha_i - \kappa - \frac{\Gamma_i((n-1)\alpha_j - \alpha_i - \kappa, \alpha_i, \alpha_{-i})}{\xi}\right) + \mathbb{P}(N_i \leq (n-1)\alpha_j - \kappa)$$
$$\left(1 - \mathbb{P}\left(R_i - D_i - Q \leq (n-1)\alpha_j - \alpha_i - \kappa - \frac{\Gamma_i((n-1)\alpha_j - \alpha_i - \kappa, \alpha_i, \alpha_{-i})}{\xi}\right)\right)\ dF_{\tilde{K}}(\kappa)$$
$$\geq \int_0^{(n-1)\alpha_i} \mathbb{P}\left(R_j - D_j - Q \leq (n-1)\alpha_i - \alpha_j - \kappa - \frac{\Gamma_j((n-1)\alpha_i - \alpha_j - \kappa, \alpha_j, \alpha_{-j})}{\xi}\right) + \mathbb{P}(N_j \leq (n-1)\alpha_i - \kappa)$$
$$\left(1 - \mathbb{P}\left(R_j - D_j - Q \leq (n-1)\alpha_i - \alpha_j - \kappa - \frac{\Gamma_j((n-1)\alpha_i - \alpha_j - \kappa, \alpha_j, \alpha_{-j})}{\xi}\right)\right)\ dF_{\tilde{K}}(\kappa) = \frac{c}{\psi},$$

a contradiction. Thus, $\alpha_i = \alpha_j$. Let $\alpha_i = \alpha_j = \alpha$. This also implies that $\Gamma_i = \Gamma_j = \Gamma$ for some function Γ. Also, Γ and α is jointly determined by Eq. (19).

Lemma 29. *There is at least one pair of α and Γ that satisfies Eq. (19).*

The proof is omitted for it is similar to the proof of Lemma 20.

G Internet Appendix: Additional proofs

G.1 Proof of Lemma 19

Fix any $\alpha_i, \alpha_j \in [N_{min}, \infty)$ such that $\alpha_i < \alpha_j$. Assume bank i pays $a_i = \min((L_i + \alpha_i)^+, N_i)$ and bank j pays $a_j = \min((L_j + \alpha_j)^+, N_j)$. Banks optimize in the trading game so Lemma 4 holds.

Lemma 30. *Fix any $x \geq y$ and $x > N_{min}$. If $y \leq N_{min}$ then $\mathbb{P}(a_i \leq x) \geq \mathbb{P}(a_j \leq y)$. If $y > N_{min}$ and $\forall n \in [N_{min}, y)$, $\mathbb{P}(a_i \leq n - (y - \alpha_j)) \geq \mathbb{P}(a_j \leq n - (x - \alpha_i))$, then $\mathbb{P}(a_i \leq x) \geq \mathbb{P}(a_j \leq y)$.*

Proof. By Lemma 9, in equilibrium $\psi\mathbb{P}(a_j \leq \alpha_i) \geq c$ and for any $\vartheta < \alpha_i$, $\psi\mathbb{P}(a_j \leq \vartheta) \leq c$. When $y \leq -\alpha_j$, then by Lemma 2,

$$\begin{aligned}\Gamma_j(y - \alpha_j, \alpha_j, \alpha_i) =& \psi\mathbb{P}(a_i \leq -y) \geq c \\ \geq& \int_{n \in [N_{min}, x)} \psi\mathbb{P}(a_j \leq n - (x - \alpha_i))\, dF_N(n) + c\mathbb{P}(N_i \geq x) = \Gamma_j(x - \alpha_i, \alpha_i, \alpha_j)\end{aligned}$$

When $-\alpha_j < y \leq N_{min}$, then by Lemma 2,

$$\Gamma_j(y - \alpha_j, \alpha_j, \alpha_i) = c \geq \int_{n \in [N_{min}, x)} \psi\mathbb{P}(a_j \leq n - (x - \alpha_i))\, dF_N(n) + c\mathbb{P}(N_i \geq x) = \Gamma_i(x - \alpha_i, \alpha_i, \alpha_j).$$

When $y > N_{min}$ and $\forall n \in [N_{min}, y)$,

$$\mathbb{P}(a_i \leq n - (y - \alpha_j)) \geq \mathbb{P}(a_j \leq n - (x - \alpha_i)).$$

Then

$$\int_{n \in [N_{min}, y)} \psi\mathbb{P}(a_i \leq n - (y - \alpha_j))\, dF_N(n) \geq \int_{n \in [N_{min}, y)} \psi\mathbb{P}(a_j \leq n - (x - \alpha_i))\, dF_N(n).$$

Therefore, by Lemma 2

$$\begin{aligned}\Gamma_j(y - \alpha_j, \alpha_j, \alpha_i) =& \int_{n \in [N_{min}, y)} \psi\mathbb{P}(a_i \leq n - (y - \alpha_j))\, dF_N(n) + c\mathbb{P}(N_i \geq y) \\ \geq& \int_{n \in [N_{min}, x)} \psi\mathbb{P}(a_j \leq n - (x - \alpha_i))\, dF_N(n) + c\mathbb{P}(N_i \geq x) = \Gamma_i(x - \alpha_i, \alpha_i, \alpha_j)\end{aligned}$$

In any case,

$$x - \alpha_i - \frac{\Gamma_i(x - \alpha_i, \alpha_i, \alpha_j)}{\xi} \geq y - \alpha_j - \frac{\Gamma_j(y - \alpha_j, \alpha_j, \alpha_i)}{\xi}.$$

Since $R_i - D_i$ and $R_j - D_j$ have the same distribution,

$$\mathbb{P}\left(R_i - D_i - Q + \lambda^{-1} \leq x - \alpha_i - \frac{\Gamma_i(x-\alpha_i, \alpha_i, \alpha_j)}{\xi}\right) \geq \mathbb{P}\left(R_j - D_j - Q + \lambda^{-1} \leq y - \alpha_j - \frac{\Gamma_j(y-\alpha_j, \alpha_j, \alpha_i)}{\xi}\right).$$

Lemma 4 and this inequality imply that $\mathbb{P}(L_i \leq x - \alpha_i) \geq \mathbb{P}(L_j \leq y - \alpha_j)$. Since $\mathbb{P}(N_i \leq x) \geq \mathbb{P}(N_j \leq y)$,

$$\begin{aligned}
\mathbb{P}(a_i \leq x) =& \mathbb{P}(L_i + \alpha_i \leq x) + \mathbb{P}(N_i \leq x) - \mathbb{P}(N_i \leq x)\mathbb{P}(L_i + \alpha_i \leq x) \\
\geq& \mathbb{P}(L_i + \alpha_i \leq x) + \mathbb{P}(N_i \leq y) - \mathbb{P}(N_i \leq y)\mathbb{P}(L_i + \alpha_i \leq x) \\
=& \mathbb{P}(L_i + \alpha_i \leq x) + \mathbb{P}(N_j \leq y) - \mathbb{P}(N_j \leq y)\mathbb{P}(L_i + \alpha_i \leq x) \\
\geq& \mathbb{P}(L_j + \alpha_j \leq y) + \mathbb{P}(N_j \leq y) - \mathbb{P}(N_j \leq y)\mathbb{P}(L_j + \alpha_j \leq y) \\
=& \mathbb{P}(a_j \leq y).
\end{aligned}$$

□

Define set $\mathcal{I}_0 = [N_{min}, \alpha_i)$, and functions $A_0, B_0 : \mathcal{I}_0 \to \mathbb{R}$ such that $A_0(n_0) = n_0$, and $B_0(n_0) = n_0$. Define correspondence $\mathcal{I}_1(n) = [N_{min}, B_0(n))$, and functions $A_1(n_0, n_1) = n_1 - B_0(n_0) + \alpha_j$ and $B_1(n_0, n_1) = n_1 - A_0(n_0) + \alpha_i$ on domain $\mathcal{D}_1 = \{(n_0, n_1) \mid n_0 \in \mathcal{I}_0, n_1 \in \mathcal{I}_1(n_0)\}$. Iteratively, given $\ell \in \mathbb{N}_+$, $\mathcal{I}_\ell(n_0, n_1, \cdots, n_{\ell-1})$, $A_\ell(n_0, n_1, \cdots, n_\ell)$ and $B_\ell(n_0, n_1, \cdots, n_\ell)$, define

$$\mathcal{I}_{\ell+1}(n_0, n_1, \cdots, n_\ell) = [N_{min}, B_\ell(n_0, n_1, \cdots, n_\ell)).$$

Then define functions $A_{\ell+1}(n_0, n_1, \cdots, n_\ell, n_{\ell+1})$, $B_{\ell+1}(n_0, n_1, \cdots, n_\ell, n_{\ell+1})$ on domain $\mathcal{D}_{\ell+1} = \{(n_0, n_1, \cdots, n_\ell, n_{\ell+1}) \mid n_0 \in \mathcal{I}_0, n_1 \in \mathcal{I}_1(n_0), \cdots, n_\ell \in \mathcal{I}_\ell(n_0, n_1, \cdots, n_{\ell-1}), n_{\ell+1} \in \mathcal{I}_{\ell+1}(n_0, n_1, \cdots, n_\ell)\}$ such that

$$\begin{aligned}
A_{\ell+1}(n_0, n_1, \cdots, n_\ell, n_{\ell+1}) &= n_{\ell+1} - B_\ell + \alpha_j \\
B_{\ell+1}(n_0, n_1, \cdots, n_\ell, n_{\ell+1}) &= n_{\ell+1} - A_\ell + \alpha_i.
\end{aligned}$$

It is easy to check that

$$A_\ell(n_0, n_1, \cdots, n_\ell) - B_\ell(n_0, n_1, \cdots, n_\ell) = \ell(\alpha_j - \alpha_i).$$

In addition, $B_\ell(n_0, n_1, \cdots, n_\ell) \leq \alpha_i$ and $A_\ell(n_0, n_1, \cdots, n_\ell) \geq N_{min}$ for all ℓ on their domain $\mathcal{D}_\ell$.

Because $A_\ell(n_0, n_1, \cdots, n_\ell) - B_\ell(n_0, n_1, \cdots, n_\ell) \to \infty$ as $\ell \to \infty$, there is one $T \in \mathbb{N}_+$ such that

$$\mathbb{P}(a_i \leq A_T(n_0, n_1, \cdots, n_T)) \geq \mathbb{P}(a_j \leq B_T(n_0, n_1, \cdots, n_T))$$

on domain $\mathcal{D}_T$. Then by Lemma 30

$$\mathbb{P}(a_i \leq A_{T-1}(n_0, n_1, \cdots, n_{T-1})) \geq \mathbb{P}(a_j \leq B_{T-1}(n_0, n_1, \cdots, n_{T-1}))$$

on domain $\mathcal{D}_{T-1}$. Apply Lemma 30 repeatedly, we will arrive

$$\mathbb{P}(a_i \le A_0(n_0)) = \mathbb{P}(a_i \le n_0) \ge \mathbb{P}(a_j \le B_0(n_0)) = \mathbb{P}(a_j \le n_0)$$

for all $n_0 \in [N_{min}, \alpha_i)$.

G.2 Proof of Lemma 20

Let $\mathcal{L}$ be the set of continuous decreasing functions defined on $(-\infty, 0]$.

$$\mathcal{L} = \{f : (-\infty, 0] \to [0, c] \;\big|\; f(x) \text{ is decreasing and continuous. }\}$$

Consider a map $\mathcal{T} : \mathcal{L} \to \mathcal{L}$ such that for any $\Gamma_0 \in \mathcal{L}$, we have $\mathcal{T}(\Gamma_0) = \Gamma$. Γ is characterized by a constant α, defined as

$$\alpha = \inf \left\{ \vartheta \ge N_{min},\; F_R\left(-\frac{\Gamma_0(0)}{\xi}\right) + F_N(\vartheta)\left(1 - F_R\left(-\frac{\Gamma_0(0)}{\xi}\right)\right) \ge \frac{c}{\psi} \right\},$$

in the following way: for all $-\alpha + N_{min} \le y \le 0$,

$$\begin{aligned}\Gamma(y) = &\int_{n\in(N_{min},(y+\alpha))} \psi\Big(F_N(n-y) + (1 - F_N(n-y))F_R(n - y - \alpha - \frac{\Gamma_0(n-y-\alpha)}{\xi})\Big)\, f_N(n)dn \\ &+ c(1 - F_N(y+\alpha))\end{aligned}$$

and for for all $y \le -\alpha + N_{min}$, $\Gamma(y) = c$. We first show that $\mathcal{T}$ is well-defined. Since F_R is monotone, F_R only has a finite number of discontinuous points. Thus, when Γ_0 is continuous, Γ is continuous. In addition, by definition $\alpha \in [N_{min}, \hat{N}]$. Note that by construction,

$$F_R\left(-\frac{\Gamma_0(0)}{\xi}\right) + F_N(\vartheta)\left(1 - F_R\left(-\frac{\Gamma_0(0)}{\xi}\right)\right) < \frac{c}{\psi}$$

for all $\vartheta < \alpha$. Also note that when $n = y + \alpha$,

$$\psi\Big(F_N(n-y) + (1 - F_N(n-y))F_R(n - y - \alpha - \frac{\Gamma_0(n-y-\alpha)}{\xi})\Big) = c.$$

It follows that Γ decreases on $\mathbb{R}^-$. Hence $\Gamma \in \mathcal{L}$ and $\mathcal{T}$ is well-defined.

Let $\mathcal{L}$ be endowed with L^∞ norm. It follows that $\mathcal{L}$ is a closed and compact space, and $\mathcal{T}$ is continuous. Schauder fixed-point theorem implies that there exist at least one fixed point.