Treasury Bill Shortages and the Pricing of Short-Term Assets*

Adrien d'Avernas† Quentin Vandeweyer‡

March 15, 2023

Abstract

We propose a model of post-GFC money markets and monetary policy implementation. In our framework, capital regulation may deter banks from intermediating liquidity derived from holding reserves to shadow banks. Consequently, money markets can be segmented, and the scarcity of Treasury bills available to shadow banks is the main driver of short-term spreads. In this regime, open market operations have an inverse effect on net liquidity provision when swapping ample reserves for scarce T-bills or repos. Our model quantitatively accounts for post-2010 time series for repo rates, T-bill yields, and the Fed's reverse repo facility usage.

Keywords: Repo, Reverse Repo Facility, Money Markets, Shadow Banks, Monetary Policy.
JEL Classifications: E43, E44, E52, G12

*We would like to thank Nordine Abidi, Markus Brunnermeier, Nicolas Coeurdacier, Dong Beom Choi, Darrell Duffie, Wenxin Du, Stéphane Guibaud, Marie Hoerova, Arvind Krishnamurthy, Caterina Mendicino, Stefan Nagel, Monika Piazzesi, Raghuram Rajan, David Skeie, Jean-David Sigaux, and Etienne Wasmer for their valuable comments as well as participants in seminars, workshops, and conferences at the Booth School of Business, Stanford GSB, Northwestern Kellogg, Toulouse School of Economics, the FRB of New York, Olin School of Business, McGill University, the University of Luxembourg, Amsterdam VU, the Riksbank, the European Central Bank, the Bank of Canada, and the Bank of Spain. We acknowledge gracious support from the Fama-Miller Center for Research in Finance and thank Livia Amato and Baiyang Han for their excellent research assistance. Previous versions of this paper were circulated under the title "Treasury debt and the pricing of short-term assets." Adrien d'Avernas gratefully acknowledges financial support from the Nasdaq Nordic Foundation.

†Stockholm School of Economics

‡University of Chicago: Booth School of Business.

1 Introduction

Before the Great Financial Crisis (GFC), short-term rates were tightly controlled by the Federal Reserve (Fed) through variations in the supply of scarce reserves available to banks. Since then, the Fed has been implementing an "ample reserve" framework by satiating banks' demand for reserves and varying the interest it pays on these balances. Although this regime was initially believed to be stable and simple to implement, short-term money market rates such as repos, fed funds, and T-bill rates have been markedly volatile during the last decade. For instance, in the first semester of 2018, these short-term rates, puzzlingly, hiked beyond levels anticipated by the Fed and prompted doubts about its ability to control short-term rates.[1] Conversely, since January 2021, short-term rates have been trending down and eventually reached the interest rate the Fed is paying on its new reverse repo (RRP) facility, introduced to prevent money market rates from falling below that level. Volumes at this facility have since reached $2.5 trillion.

This paper proposes a framework to explain post-GFC money market rates and holdings by traditional and shadow banks. In our model, a combination of large excess reserves balances with binding regulatory capital requirements creates a market segmentation in which traditional banks cease to intermediate liquidity to shadow banks. When this phenomenon occurs, the pricing of reserves disconnects from the pricing of other money-like assets, such as T-bills and repos held by shadow banks. While reserves are in excess for traditional banks, T-bills are scarce and valuable to shadow banks. Consequently, changes in the supply of T-bills are the main driver of money market spreads and open market operations (that expand the supply of reserves but reduce the supply of T-bills or repos) reduce liquidity and increase liquidity risk.

Our research is motivated by a series of recent developments in US dollar money markets. As shown by Andersen, Duffie, and Song (2019), non-risk-weighted leverage ratios limit financial intermediaries' abilities to arbitrage small spreads between markets.[2] For instance, Du, Tepper, and Verdelhan (2018) document sustained deviations from covered interest rate parity since 2008, and Duffie and Krishnamurthy (2016) highlight the impact of the introduction of stricter constraints on money market spreads. In line with these developments, we further document in Figure 1 a strong relationship between the supply of T-bills (adjusted for Fed and government-

[1] During a press conference on June 13, 2018, Chairman Powell answered a question related to the recent hike in money market rates as: "[W]e're looking carefully at that and, you know, the truth is we don't know with any precision. Really, no one does. [...] I think there's a lot of probability on the idea of just high bill supply leads to higher repo costs, higher money market rates generally, and the arbitrage pulls up federal funds rate towards [interest on reserves]. We don't know that that's the only effect." Examples of this negative perception of the ability of the Fed to control short-term interest rates can be found in the financial press, as illustrated by the article by Alex Harris from May 30, 2018, titled "As Fed Loses Control of Overnight Rates, Blame Shifts to T-Bills" (Bloomberg, https://www.bloomberg.com/opinion/articles/2018-01-16/the-fed-is-losing-control-of-the-financial-markets, accessed on the 01/08/2019).

[2] Leverage ratio regulation was first introduced in the US in the 1980s. Initially conceptualized as a last-resort constraint to complement risk-weighted regulation, it was designed not to bind for most institutions (Haubrich, 2020). However, following the 2008 financial crisis, these leverage ratios became increasingly binding with the subsequent introduction of Basel II in April 2008; the Dodd-Frank Act in July 2010; the revision of FDIC Assessment Methodology in April 2011; Basel III phase-in, with a minimum leverage ratio of 3% since 2013; and a supplementary leverage ratio (SLR) of an additional 2% for top-tier bank holding companies in January 2014.

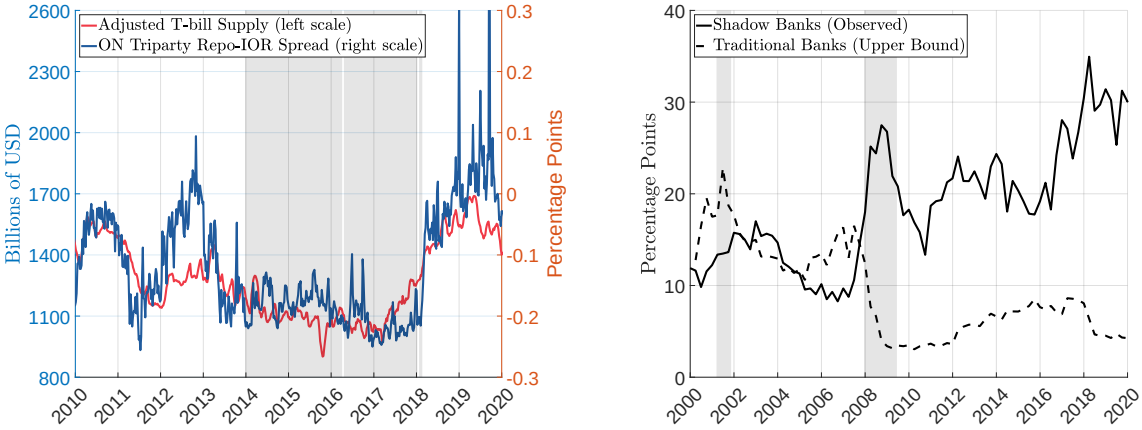


Figure 1: Share of Available T-bills Held by Shadow and Traditional Banks. The left panel displays the quantity of T-bill outstanding adjusted by removing Fed holdings and government-only fund holdings (blue line, left scale) and the spread between overnight triparty repo rates and the interest on reserves. The shaded area represents periods where volumes at the reverse repo facility exceed \$15 billion. The right panel displays the ratio of the value of T-bills held by shadow banks to the stock of T-bills available and the ratio of an upper bound measure of the share of T-bills held by traditional banks to the stock of T-bills available. Shadow banks' holdings are computed as the sum of holdings from money funds and mutual funds from flow of funds data. The stock of available T-bills is computed as the total amount outstanding minus the holdings of the Fed from flows of funds data. The upper bound measure for traditional banks is computed as the total value of assets with a maturity of less than a year in depository institutions. Data sources: Bloomberg, Call Reports, Federal Reserve Bank of New York, FRED, Treasury Direct.

only money funds holdings) and triparty repo spreads to interest on reserves and an increase in the proportion of T-bills held by shadow banks relative to banks.

To account for these developments, we model heterogeneous banks supplying liquid deposits to households while subject to a liquidity management problem and a balance sheet cost. Liquid assets carry a liquidity premium as holding them helps to avoid costly fire sales. The banking sector is composed of traditional and shadow banks. While traditional banks can hold and trade central bank reserves and T-bills, shadow banks are limited to T-bills. This assumption aligns with the institutional restriction that only financial firms carrying a banking license are authorized to have an account at the Fed. This rule excludes institutions usually associated with the liquidity management of shadow banks that are active in the repo market—such as money market funds, securities lenders, and hedge funds—from holding reserves at the central bank. Moreover, traditional banks are assumed to be subject to a cost when increasing the size of their balance sheets. This balance sheet cost is motivated as originating from a regulatory leverage ratio that forces banks to finance arbitrage positions with equity and, thereby, generates a debt-overhang friction à la Myers (1977) and Andersen, Duffie, and Song (2019). A treasury and a central bank complete the model by supplying scarce public liquid assets to banks and influencing short-term rates.

This paper's first contribution is to provide a tractable model to account for pre- and post-GFC monetary policy implementation regimes. Before the crisis, the Fed was not authorized to

pay interest on reserves, and monetary policy was implemented through open market operations, adjusting the supply of liquid assets available to banks. Since the crisis, the Fed has used the interest on reserves as its main policy tool to lift rates while maintaining a large balance sheet, complemented with the introduction of a reverse repo facility for money market funds in 2013.[3] The key feature of our model that allows us to capture the two regimes is the ability of the central bank to control both the liquidity premium on reserves—through the supply of liquid assets available to banks—and the interest it pays on reserves. Since both of these variables affect short-term nominal rates, there is one degree of freedom in the implementation problem of monetary policy. This result provides a general equilibrium extension of Poole (1968) and holds with and without nominal frictions. With this feature, the model allows us to compare the pre-2008 period when the Fed did not pay interest on reserves and the post-2010 period when the liquidity premium on reserves was nil due to the large amounts of excess reserves outstanding.

The paper's second contribution is to show that the combination of a large amount of excess reserves and a binding regulatory leverage ratio creates a segmentation of money markets in which open market and repo operations from the central bank have a reverse effect on the net supply of liquid assets. In this regime, traditional banks are fully satiated with liquidity because the supply of reserves is large, but they are unwilling to intermediate this liquidity to shadow banks.[4] For this reason, while the liquidity premium on reserves drops to zero, the liquidity premium of other money market assets, such as T-bills, remains positive because liquid assets are still scarce for shadow banks. This outcome has the consequence whereby yields on all non-reserves liquid assets have to drop when the supply of T-bills becomes scarcer. This mechanism would explain why most money market rates have been trading below the interest paid on reserves in various countries in which traditional banks hold large amounts of excess reserves (Arrata, Nguyen, Rahmouni-Rousseau, and Vari, 2017; Duffie and Krishnamurthy, 2016). The model, therefore, also offers a natural interpretation of the Fed's RRP as an instrument through which the central bank creates liquid assets that shadow banks can directly hold and prevents interest rates from falling below the interest paid at this facility. Interestingly, this interpretation is opposite to a common view of the facility as "liquidity absorbing."[5] The reason for this divergence in interpretation is that in our model, even though the facility removes most liquid central bank reserves to create less liquid repos, it does so at a time when the marginal value of liquidity from reserves for banks is below the marginal value of repos for shadow banks. Hence, the net impact of the facility on aggregate liquidity is positive. A similarly inverse response occurs when the central bank increases reserves through open market operations in

[3]In addition, to prevent short-term rates from "leaking" too far below the interest on reserves, in 2013, the Fed opened a reverse repo (RRP) facility that enables money market funds to deposit funds with the Fed for a given policy rate called the RRP rate. See Amstad and Martin (2011) for a detailed account of various pre- and post-GFC practices.

[4]When banks borrow from institutions, such as money market funds, in repo markets, they are effectively creating a safe overnight asset akin to a deposit in which these institutions can "park" their liquid balances while earning interest.

[5]See, for instance, https://www.reuters.com/article/us-usa-bonds-repo-explainer/explainer-u-s-repo-market-flirts-with-negative-rates-as-fed-seeks-to-absorb-excess-cash-idUSKBN2C32AI, accessed on 12/22/2021.

the segmented regime.

A third contribution of the paper is to explain why the supply of T-bills has had an outsized effect on repo yields since the GFC. Our model implies that without a leverage ratio constraint, money markets are fully integrated because traditional banks can expand their balance sheets to intermediate liquidity through repo without any cost. In this regime, all liquid assets are priced by a single factor controlled by the Fed to target short-term nominal rates. Consequently, new issuance of T-bills from the Treasury must be sterilized by the Fed through offsetting open market operations since it would otherwise put downward pressure on this target rate. Since money market assets are perfect substitutes for each other, these offsetting operations fully neutralize the effect of a change in the supply of T-bills *on all liquid assets*. This result is consistent with the findings of Nagel (2016), showing that T-bill supply loses all explanatory power on liquidity premia when controlling for monetary policy proxied by the fed funds rate. In contrast, in an economy with a regulatory leverage ratio, banks internalize the cost of their balance sheet so that it is sometimes unprofitable to intermediate liquidity to shadow banks in repos. In this inaction region, the pricing of reserves is disconnected from the pricing of T-bills. As a consequence, variations in the supply of T-bills only affect the liquidity premium of liquid assets held by shadow banks, and the central bank does not conduct offsetting open market operations. Hence, yields on both repo and T-bills are free to react to changes in the supply of T-bills. The model further predicts that money market rates react to exogenous changes in the supply of T-bills if and only if the repo rate is above the RRP rate. Once this rate is reached, variations in the supply of T-bills show up as adjustments in the quantities of liquid assets created by the Fed within the facility.

In the last section of this article, we quantify the main equations of our model ruling the dynamics of T-bill and repo spreads as well as volumes at the RRP facility. We first estimate each equation individually through multiple specifications, including controls and instruments for T-bill supply and the government-only money market funds inflows. We find that estimated parameters are consistent with the predictions of the model with an increase in T-bill supply of a $100 billion leading to an increase in T-bill rate of 4 basis points or a decrease in RRP volumes of $12 billion. We then proceed to estimate the same equations while considering model-implied cross-equation restrictions and find that the model yields larger estimates without significantly affecting the ability of the model to fit observations.

In particular, the model correctly predicts the increase in money market rates observed at the beginning of 2018, as well as the subsequent drop in usage of the RRP facility. This event is interpreted through the lens of the model as follows: The increase in the supply of T-bills drives the liquidity premia on various money market instruments down. As the rate on illiquid assets is pinned down by the interest on reserves, this narrowing of the spread takes the form of an increase in money market rates, which grows closer to the interest on reserves. In the reverse direction, the model also accurately predicts the movement in repo and T-bill rates since the Covid shock of March 2020. Initially, as a reaction to the large T-bill issuance by the Treasury, repo and T-bill rates traded very close to the interest on reserves, reflecting their abundance.

Subsequently, these spreads started to widen around January 2021, and both rates dropped to the RRP rate when the Treasury started converting T-bills into longer-term debt. Further reductions in the supply of T-bills were absorbed by RRP volumes, which had risen to $1.5 trillion by October 2021 to reflect a doubling in shadow bank demand for liquidity since March 2020.

We also further develop an extension with elastic deposit demand to account for the effect of monetary policy on deposit market dynamics (Drechsler, Savov, and Schnabl, 2017). In particular, as shown by Xiao (2020), an increase in the Fed target interest rate leads to a reallocation of deposits from traditional banks to shadow banks, which attracts attentive yield-sensitive investors, while traditional banks increase their interest margin by keeping their deposit rate close to zero. In our model, such a reallocation of deposits from traditional to shadow bank generates an increase in the demand for T-bills relative to reserves and is able to aggravate its scarcity. We find this mechanism to be quantitatively mostly muted in most of the 2010-2022 period. We then use these estimates to carry a counterfactual exercise which informs us that without this facility or a significant increase in banks' balance sheet capacity, T-bill and repo rates would have dropped by an additional 60 bps by October 2021, which underlines how central the T-bill supply and the RRP facility have become to US monetary policy implementation.

Related Literature A large literature examines liquidity premia on short-term assets. In particular, the idea of a convenience yield in government bonds, as in money, is found in Patinkin (1956); Tobin (1963); Bansal and Coleman (1996); Duffee (1996); Krishnamurthy and Vissing-Jorgensen (2012). Specifically, Greenwood, Hanson, and Stein (2015) document that T-bill supply has a crowding out effect on private liquidity transformation while Nagel (2016) finds that when controlling for monetary policy, T-bills and reserves appear to be perfect substitutes. In recent work, Krishnamurthy and Li (forthcoming) find that when considering nonlinearities, these appear to be strong but imperfect substitutes. In addition, government bonds used as an imperfect means of payment are an important feature of the neo-monetarist literature with trade frictions (Andolfatto and Williamson, 2015; Venkateswaran and Wright, 2013). Bech, Klee, and Stebunovs (2011) and Lenel, Piazzesi, and Schneider (2019) explain the convenience yields on short-term bonds as originating from intermediaries' demand for collateral to back inside money. In a closely related work, Duffie and Krishnamurthy (2016) link the increased dispersion in money market spreads and lower monetary policy pass-through to the introduction of SLR and propose a search friction model of money market in which the supply of T-bills can affect money market spreads and reverse repo take up through a rationing mechanism. In contemporaneous work, Martin, McAndrews, Palida, and Skeie (2019) propose a model with convex balance sheet cost in which the RRP facility absorbs excessive reserves while the T-bill supply competes with the RRP. Huber Wang (2023) proposes a structural estimation of dealers' market power over money funds in Triparty repo markets. This paper adds to this literature by stressing and quantifying the increasing importance of shadow banks' demand and T-bills supply in driving short-term rates in a comprehensive model in which liquid assets bear

a premium as a hedge against liquidity risk.

This work builds on the macro-finance literature with a financial sector (Brunnermeier and Sannikov, 2014; He and Krishnamurthy, 2013), and shares with these articles an incomplete market structure. As in Brunnermeier and Sannikov (2016), the model features both inside and outside money while as in Drechsler, Savov, and Schnabl (2018), funding liquidity shocks may affect risk premia and asset prices through the balance sheets of intermediaries. In the banking literature, Holmstrom and Tirole (1998) and Diamond and Rajan (2001, 2005) characterize optimal liquidity provision when interbank markets are affected by liquidity shocks. Afonso and Lagos (2015) and Bech and Monnet (2016) develop over-the-counter models of the interbank market with random matching to understand its trading dynamics. Close to our paper, Bianchi and Bigio (2022); Piazzesi and Schneider (2021); and De Fiore, Hoerova, and Uhlig (2018) include interbank markets in macroeconomic models and study the effect of lender-of-last-resort monetary policy on macroeconomic variables. This paper adds to this literature by introducing a micro-founded liquidity management problem to an asset-pricing model in which a subset of banks does not have direct access to central bank reserves. In this regard, it also relates to the literature on limited arbitrage following Gromb and Vayanos (2002). In particular, our paper broadens the fed funds market segmentation result of Bech and Klee (2011) to include repos and T-bills and account for endogenous policy from the Treasury and the Fed.

This paper is also linked to the literature on shadow banking and the shortage of safe assets. The demand for safe assets from shadow banks and the role of money market instruments in filling this gap and creating financial fragility are studied by Stein (2012); Caballero (2006); Lenel (2020); Sunderam (2015); and Li (2022). Moreover, Plantin (2015); Huang (2018); and Ordoñez (2018) study the emergence of the shadow banking sector as a consequence of regulatory arbitrage. Infante (2020) also documents that T-bills and repo are substitutes, while the longer-term Treasury bond supply has an ambiguous relationship with repos. This paper contributes to the literature by investigating the implications of a market segmentation appearing when it is costly for traditional banks to intermediate liquidity to shadow banks. In this regard, our paper relates to a large literature documenting the emergence of arbitrage opportunities in post-GFC financial markets (Avdjiev, Du, Koch, and Shin, 2019; Boyarchenko, Costello, and Shachar, 2020; Du, Tepper, and Verdelhan, 2018; Siriwardane, Sunderam, and Wallen, 2022). This paper also complements d'Avernas and Vandeweyer (2021) and Copeland, Duffie, and Yang (2022), who study how money markets adjust to scarcity of reserves caused by intraday liquidity constraints and explain late-2019 and early-2020 repo and Treasury market instabilities.

2 The Model

This section presents a model of local money market segmentation in which traditional and shadow banks trade exposure to liquidity risk. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that satisfies the usual conditions and assume that all stochastic processes are adapted. The economy evolves in continuous time with $t \in [0, \infty)$ and is populated by a continuum of traditional banks, shadow

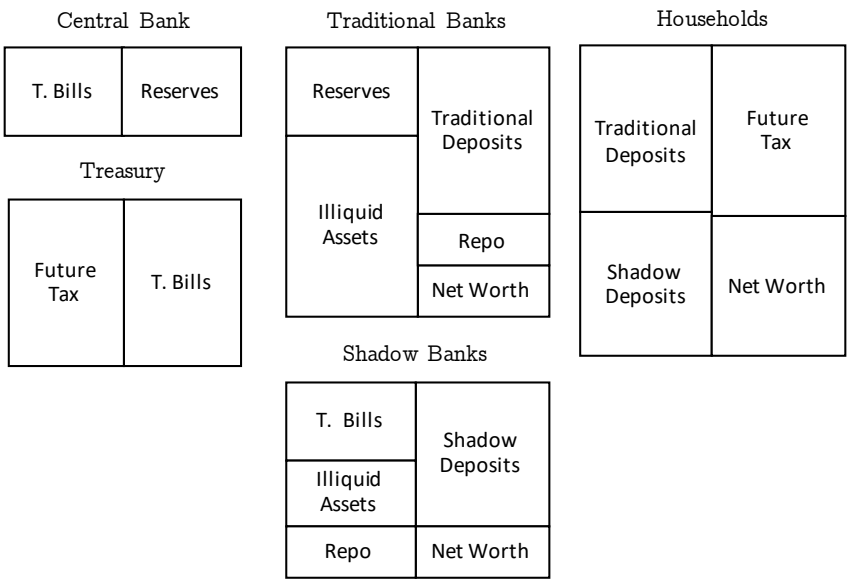


Figure 2: Chart of Sectors' Balance Sheets

banks, and households of mass one, as well as the Treasury and a central bank. Figure 2 depicts the balance sheets of the different sectors in the economy. The Treasury issues T-bills against future tax liabilities; the central bank holds some of the outstanding T-bills by issuing reserves to the banking sector; and households hold their wealth in both traditional bank deposits and shadow bank deposits. The two types of banks issue deposits to finance their holdings of illiquid assets and hold liquid assets as a buffer against deposit shocks. All rates and prices are expressed in real terms except when explicitly specified otherwise.

2.1 Environment

Preferences All agents have logarithmic preferences over their consumption rate c_t of their net worth n_t with a time preference ρ:

$$V_t = \mathbb{E}_t \left[\int_t^\infty e^{-\rho(u-t)} \log(c_u n_u) du \right].$$

Technology There is a unit of illiquid risk-free capital producing a flow of real output, y_t, with constant productivity, which, for simplicity, is assumed to be equal to the time preference parameter ρ. Thus, because all agents have logarithmic preferences, the price of capital is constant and equal to 1 and the total output is equal to ρ.[6]

[6]The total net worth in the economy is equal to $q_t k_t$ where q_t is the price of capital and $k_t = 1$ is the total stock of capital. Total output is given by $y_t = \rho k_t$. Finally, with logarithmic preferences, aggregate consumption is given by $c_t = \rho q_t k_t$. Thus, $q_t = 1$.

Nominal Definition The economy does not feature any nominal frictions, and the ability of the central bank to influence inflation is derived from the role of reserves as the unit of account. Nominal output is defined as $y_t^{\$} = y_t p_t$, where p_t is the price level. Inflation π_t is then defined by the drift of the deterministic law of motion of aggregate prices: $dp_t/p_t = \pi_t dt$. Although for expositional simplicity, we present here a model without nominal friction as in the second chapter of Woodford (2003), Appendix D extends our model to incorporate nominal frictions without affecting our main results.

Liquidity Management All banks are subject to idiosyncratic funding shocks. After a negative realization of the funding shock, some deposits in a given bank are transferred to another bank. This process can be interpreted either as a feature of normal payment flows from depositors or as an abnormal run on a given bank. This reshuffling creates a funding gap for one (the deficit bank) and a funding surplus for the other (the surplus bank). The sequence of actions takes place in a short period of time in which illiquid assets can only be traded at a loss with respect to their fundamental value.[7]

To avoid bearing the cost of these fire sales, banks can hold liquid money-like assets as a buffer against funding shocks. Two assets with this property are issued by the government: T-bills b_t and central bank reserves m_t. Importantly, reserves can only be held by traditional banks (and not shadow banks), while all types of banks can hold T-bills. In addition, traditional banks can also issue a liquid asset to shadow banks, which we call repo. This instrument can, thereby, be used to pass on the liquidity from reserves to shadow banks.

Throughout this article, we maintain the assumption that liquidity services of reserves are higher than the ones of T-bills, which are themselves higher than the ones of repos: $\theta^m > \theta^b > \theta^p$.[8] In the Online Appendix, we show that such a problem converges in continuous time to the following idiosyncratic, but not diversifiable, transfers of wealth:

$$\text{traditional banks:} \qquad \psi_t d\tilde{Z}_t = \lambda \left[\sigma^d w_t^d - \theta^m w_t^m - \theta^b w_t^b - \theta^p w_t^p\right]^+ d\tilde{Z}_t, \tag{1}$$

$$\text{shadow banks:} \qquad \overline{\psi}_t d\tilde{Z}_t = \lambda \left[\sigma^d \overline{w}_t^d - \theta^b \overline{w}_t^b - \theta^p \overline{w}_t^p\right]^+ d\tilde{Z}_t, \tag{2}$$

where the variables $w_t^d, w_t^m, w_t^b, w_t^p$ represent the portfolio weights of traditional bank holdings on deposits, reserves, T-bills, and repos, respectively, $d\tilde{Z}_t$ is a standard adapted Brownian motion, which is idiosyncratic to an individual bank, and $[x]^+$ denotes $\max\{x, 0\}$. The upper bar notation represents variables specific to shadow banks. These equations have the following interpretation. Banks always sell their most liquid asset first since it is less costly to do so. When a negative funding shock of size $\sigma^d w_t^d$ hits a bank, it has to pay a fire-sale cost λ on the amount remaining after having disbursed liquid assets in the form of reserves w_t^m, T-bills w_t^b,

[7] A possible empirical counterpart for this hypothetical illiquid rate is the "shadow rate" for short-term rates as extrapolated from the long-term yield curve as in Lenel, Piazzesi, and Schneider (2019).

[8] This first inequality is a consequence of reserves always being accepted by banks as an ultimate means of settlement. For example, reserves are transferred without delay or cost to meet a funding shock, whereas T-bills must first be sold before a debt can be settled.

and repos w_t^p. On the flip side, when a bank receives a positive shock, it has the extra resources to purchase the asset sold by the deficit bank at a discount and make a profit on the operation. The maximum operator reflects the existence of a satiation point beyond which banks do not face liquidity risk anymore.

Treasury The Treasury issues T-bills against the future tax liabilities of other agents and is responsible for administrating redistributive lump-sum tax policies. The supply of T-bills b_t follows the stochastic process:

$$db_t = \kappa(b_t^{tar} - b_t)dt + \sigma^b \sqrt{b_t} dZ_t. \tag{3}$$

The first term, b_t^{tar}, is a time-varying target level for the Treasury that is possibly dependent on other variables in the model. This would be the case if the Treasury would increase the supply of T-bills when the liquidity premium on reserves is high. The parameter κ controls the speed of convergence to this target. Through the second term, the issuance of T-bills is subject to random shocks captured by a standard adapted Brownian Motion dZ_t that is exogenous to other variables in the model. To close the model's accounting, the net present value of future tax liabilities must equal the outstanding amount of T-bills: $b_t = \tau_t n_t + \overline{\tau}_t \overline{n}_t + \tau_t^h n_t^h$, where τ_t, $\overline{\tau}_t$, and τ_t^h are the tax liabilities per unit of wealth on traditional banks, shadow banks, and households, respectively.

Central Bank The central bank controls the supply of liquidity to the banking sector by swapping reserves for T-bills (and conversely) through open market operations and paying interest on reserves. As reserves are more liquid than T-bills, the purchase of T-bills, financed by issuing new reserves, increases the effective supply of liquidity to the banking sector. In other words, the central bank decides on the stock of reserves m_t available to banks and the quantity of T-bills to be removed from the market and held by the central bank $\underline{b}_t$ subject to the balance sheet constraint:[9]

$$\underline{b}_t = m_t. \tag{4}$$

The underline notation differentiates the central bank's holdings of T-bills $\underline{b}_t$ from the T-bills issued by the Treasury b_t. For simplicity, we assume that the central bank always operates with zero net worth and instantaneously transfers all seigniorage revenues to the Treasury. Moreover, the central bank also decides on the nominal interest rate it pays on its reserves i_t^m. Hence, the set of monetary policy decisions is $\{m_t, i_t^m\}$.

2.2 Agents' Problems

Traditional Banks Traditional banks face a Merton's (1969) portfolio choice problem augmented by the liquidity management component. Traditional banks maximize their lifetime

[9] To allow for post-GFC large balance sheet and reverse repo operations, we relax this balance sheet constraint later in Section 3.4 in order to allow the central bank also to hold illiquid assets and issue repo.

expected logarithmic utility:

$$\max_{\{w_u^i, w_u^b \geq 0, w_u^m \geq 0, w_u^p, w_u^d \geq 0, c_u \geq 0\}_{u=t}^{\infty}} \mathbb{E}_t \left[\int_t^{\infty} e^{-\rho(u-t)} \log(c_u n_u) du \right], \tag{5}$$

subject to the law of motion of wealth:

$$\begin{aligned} dn_t = & \left(w_t^i r_t^i + w_t^b r_t^b + w_t^m r_t^m + w_t^p r_t^p - w_t^d r_t^d - c_t - \mu_t^\tau \right) n_t dt \\ & + \lambda \left[\sigma^d w_t^d - \theta^m w_t^m - \theta^b w_t^b - \theta^p w_t^p \right]^+ n_t d\widetilde{Z}_t \\ & - \chi \left([w_t^i]^+ + w_t^m + w_t^b + [w_t^p]^+ \right) n_t dt, \end{aligned} \tag{6}$$

and the balance sheet constraint:

$$w_t^i + w_t^b + w_t^m + w_t^p = 1 + w_t^d + \tau_t. \tag{7}$$

Traditional banks choose their portfolio weights for illiquid assets w_t^i, T-bills w_t^b, reserves w_t^m, repos w_t^p, and deposits w_t^d given their respective interest rates r_t^i, r_t^b, r_t^m, r_t^p, and, r_t^d. Traditional banks also choose their consumption rate c_t. A negative portfolio weight for an asset corresponds to a short (liability) position for a bank, while a positive weight is a long (asset) position. The portfolio weights on reserves and T-bills are subject to a nonnegativity constraint because, by definition, these assets are liabilities of the central bank and the Treasury. This constraint does not apply to illiquid assets and repos, which can be a liability for banks.[10] When holding illiquid securities, banks increase their exposure to funding risk defined by the standard adapted Brownian $\widetilde{Z}_t$, which is idiosyncratic to the individual bank. Moreover, traditional banks are subject to a flow of transfers per unit of wealth $\mu_t^\tau dt$ to the Treasury. The variable τ_t is the net present value of these transfers, as determined by the tax policy of the Treasury. Finally, according to the last term of the law of motion for wealth, banks have to pay a cost of χ to increase the size of their balance sheet. This balance sheet cost could arise either as the shadow cost of a leverage ratio constraint in a setting in which this constraint is binding (Frazzini and Pedersen, 2014) or as a debt-overhang cost when issuing additional equity in a setting in which bank debt has some credit risk (Andersen, Duffie, and Song, 2019).

Shadow Banks Shadow banks face the same problem as traditional banks, except that they cannot hold reserves and are not subject to the balance sheet constraint:

$$\max_{\{\overline{w}_u^i, \overline{w}_u^b \geq 0, \overline{w}_u^p, \overline{w}_u^d \geq 0, \overline{c}_u \geq 0\}_{u=t}^{\infty}} \mathbb{E}_t \left[\int_t^{\infty} e^{-\rho(u-t)} \log(\overline{c}_u \overline{n}_u) du \right], \tag{8}$$

[10] In practice, a bank could short a Treasury bill and thereby take a negative exposure to price movements of T-bills. Importantly for us, when doing so, the bank does not increase the net supply of liquid T-bills available to other investors unless the position is financed with a short-maturity repo. This point is illustrated in Krishnamurthy (2002), showing that on-the-run-off-the-run spreads cannot be arbitraged away by long-short positions as higher convenience yields on on-the-run Treasuries translate in those collateral trading specials.

subject to the law of motion of wealth:

$$d\overline{n}_t = \left(\overline{w}_t^i r_t^i + \overline{w}_t^b r_t^b + \overline{w}_t^p r_t^p - \overline{w}_t^d \overline{r}_t^d - \overline{c}_t - \overline{\mu}_t^\tau\right) \overline{n}_t dt$$
$$+ \lambda \left[\sigma^d \overline{w}_t^d - \theta^b \overline{w}_t^b - \theta^p \overline{w}_t^p\right]^+ \overline{n}_t d\widetilde{Z}_t, \tag{9}$$

and the updated balance sheet constraint:

$$\overline{w}_t^i + \overline{w}_t^b + \overline{w}_t^p = 1 + \overline{w}_t^d + \overline{\tau}_t. \tag{10}$$

The interpretation of these variables—overlined to denote shadow banks—is similar to those of traditional banks. Shadow and traditional bank deposits are assumed to be perfectly nonsubstitutable.[11] Thus, the interest rate on traditional bank deposits r_t^d can differ from the interest rate on shadow bank deposits $\overline{r}_t^d$.

Households Households maximize their lifetime utility function subject to the additional assumption that they can only invest in shadow and traditional bank deposits:

$$\max_{\{c_u^h\}_{u=t}^\infty} E_t \left[\int_t^\infty e^{-\rho(u-t)} \log(c_u^h n_u^h) du\right], \tag{11}$$

subject to the law of motion of wealth:

$$dn_t^h = \left(\gamma r_t^d + (1-\gamma)\overline{r}_t^d\right)(n_t^h + \tau_t^h)dt - (c_t^h + \mu_t^{\tau,h}) n_t^h dt. \tag{12}$$

Shadow bank and traditional bank deposits are assumed to be nonsubstitutable and held in fixed proportion γ for traditional and $1-\gamma$ for shadow deposits. Hence, households merely decide on their consumption rate c^h. We make these assumptions to simplify the exposition of our main qualitative results and then relax them later when estimating the model empirically. Variables indexed by h refer to households.

Treasury Budget Constraint The budget constraint for the Treasury is given by

$$r_t^b b_t = \mu_t^\tau n_t + \overline{\mu}_t^\tau \overline{n}_t + \mu_t^{\tau,h} n_t^h + (r_t^b - r_t^m) m_t + \chi\left([w_t^i]^+ + w_t^m + w_t^b + [w_t^p]^+\right) n_t. \tag{13}$$

To pay interest on T-bills, the Treasury collects taxes from traditional banks, shadow banks, and households and seigniorage revenues from the central bank. We assume that balance sheet costs are paid as a transfer so that the real resource cost of the balance sheet constraint does not impact the price of capital in equilibrium.

[11] We relax this assumption in Section 4.4.

2.3 Solving

We provide a definition for the Markov equilibrium, make further assumptions to restrict the set of equilibria we are interested in, and derive first-order conditions.

Equilibrium We define a recursive Markov equilibrium as follows.

Definition 1. *A **Markov equilibrium** $\mathcal{M}$ in $x_t = (n_t, \overline{n}_t, b_t)$ is a set of functions $g_t = g(x_t)$ for (i) interest rates $\{r_t^i, r_t^b, r_t^m, r_t^p, r_t^d, \overline{r}_t^d\}$; (ii) individual controls for traditional banks $\{w_t^i, w_t^b, w_t^m, w_t^p, w_t^d, c_t\}$, shadow banks $\{\overline{w}_t^i, \overline{w}_t^b, \overline{w}_t^p, \overline{w}_t^d, \overline{c}_t\}$, and households $\{c_t^h\}$; (iii) monetary policy functions $\{m_t, i_t^m\}$; and (iv) transfer rules $\{\mu_t^\tau, \overline{\mu}_t^\tau, \mu_t^{\tau,h}\}$ and tax liabilities $\{\tau_t, \overline{\tau}_t, \tau_t^h\}$ such that:*

1. *Agents' optimal controls (ii) solve their respective problems given prices (i), monetary policy (iii), and transfer rules (iv).*
2. *Markets clear:*

$$\begin{aligned}
&\textit{(a) T-bills:} && w_t^b n_t + \overline{w}_t^b \overline{n}_t + \underline{b}_t = b_t, \\
&\textit{(b) reserves:} && w_t^m n_t = m_t, \\
&\textit{(c) repos:} && w_t^p n_t + \overline{w}_t^p \overline{n}_t = 0, \\
&\textit{(d) traditional deposits:} && w_t^d n_t = \gamma(1 - n_t - \overline{n}_t + \tau_t^h), \\
&\textit{(e) shadow deposits:} && \overline{w}_t^d \overline{n}_t = (1-\gamma)(1 - n_t - \overline{n}_t + \tau_t^h), \\
&\textit{(f) output:} && c_t n_t + \overline{c}_t \overline{n}_t + c_t^h (1 - n_t - \overline{n}_t) = \rho.
\end{aligned} \tag{14}$$

3. *Monetary policy variables $\{m_t, i_t^m\}$ are set as functions of the state variables only.*
4. *Transfer rules $\{\mu_t^\tau, \overline{\mu}_t^\tau, \mu_t^{\tau,h}\}$ and tax liabilities $\{\tau_t, \overline{\tau}_t, \tau_t^h\}$ are set as functions of the state variables only and satisfy the budget and the balance sheet constraints of the Treasury.*
5. *The laws of motion for the state variables in $x_t = \{n_t, \overline{n}_t, b_t\}$ are consistent with transfer and monetary policy rules.*

Thanks to logarithmic preferences, all agents of the same type choose the same set of control variables when stated as a proportion of their net worth, irrespective of the anticipated future realizations of the state variables. Hence we only have to track the supply of T-bills and the distribution of wealth between types and not within types at a given point in time. Accordingly, we characterize an equilibrium as a function of the state variables at a point in time and perform a comparative static analysis.

Equilibrium Restrictions To simplify exposition, we limit our analysis to empirically relevant cases. First, we only consider equilibria that satisfy the following condition:

$$\gamma \leq \frac{n}{n+\overline{n}}, \tag{C}$$

and thereby discard from our analysis equilibria in which the fraction of deposits in the shadow banking sector is too small compared with the relative wealth of the traditional banks. In such an equilibrium, the quantity of liquidity risk in the shadow banking sector might be so low that it is not optimal for shadow banks to hold any T-bills. Second, we restrict the analysis to equilibria in which the traditional banking sector might be a net issuer of liquid assets to the shadow banking sector: $w_t^i \geq 0$ and $w_t^p \leq 0$. Other cases would not be consistent with T-bill or repo rates lower than the IOR. Finally, we assume that both T-bills and central bank reserves are in strictly positive supply.

First-order Conditions Applying the maximum principle, we derive the first-order conditions for the three types of agents.

Traditional banks:

$$c_t = \rho, \tag{15}$$

$$r_t^i - r_t^b \geq \lambda\theta^b\psi_t \qquad \text{with equality if } w_t^b > 0, \tag{16}$$

$$r_t^i - r_t^m \geq \lambda\theta^m\psi_t \qquad \text{with equality if } w_t^m > 0, \tag{17}$$

$$r_t^i - r_t^p \leq \lambda\theta^p\psi_t + \chi \qquad \text{with equality if } w_t^p < 0, \tag{18}$$

$$r_t^i - r_t^d \leq \lambda\sigma^d\psi_t + \chi \qquad \text{with equality if } w_t^d > 0. \tag{19}$$

Shadow banks:

$$\overline{c}_t = \rho, \tag{20}$$

$$r_t^i - r_t^b \geq \lambda\theta^b\overline{\psi}_t \qquad \text{with equality if } \overline{w}_t^b > 0, \tag{21}$$

$$r_t^i - r_t^p = \lambda\theta^p\overline{\psi}_t, \tag{22}$$

$$r_t^i - \overline{r}_t^d \leq \lambda\sigma^d\overline{\psi}_t \qquad \text{with equality if } \overline{w}_t^d > 0. \tag{23}$$

Households:

$$c_t^h = \rho. \tag{24}$$

With logarithmic preferences, every agent always consumes a fixed proportion ρ. In equations (19) and (23), banks equalize the marginal benefits of issuing deposits (its liquidity risk premium) to its marginal cost (the marginal increase in liquidity risk). The first-order conditions for T-bills and reserves, given in equations (16), (17), and (21), have a similar structure but an inverse interpretation. The marginal cost is the forgone interest of holding a unit of liquid

asset—the liquidity premium—while the marginal benefit is the marginal reduction in liquidity risk. Since we are expressing spreads with respect to the illiquid asset i, the cost of balance-sheet space increases the marginal cost of liabilities. Hence, the marginal benefit of reducing a short repo position for banks also includes the additional gain of reducing the balance sheet cost by χ. Because shadow banks do not face any balance-sheet cost when holding or issuing the repo asset, they are always a marginal investor in the market for this asset such that equation (22) holds with equality.

3 Theoretical Analysis

This section presents the main theoretical results of the paper. When money markets are integrated, marginal exposure to liquidity risk is equalized between the two banking sectors. When money markets are segmented, the liquidity risks of the two banking sectors disconnect. This case arises when the reserves supply is high, and the T-bills supply is low, so traditional banks' balance sheet is too expensive for liquidity intermediation to be economical. As a consequence, the liquidity premium on T-bills responds to a change in the supply of T-bills in segmented market equilibria.

State Space Partitioning We define four sets of equilibria corresponding to changes in pricing dynamics.

Definition 2. *Let $\mathcal{I}$ be the set of* ***integrated*** *money markets equilibria, defined as* $\{\mathcal{M}(x) \in \mathcal{I} \mid w^b(x) > 0\}$.

Definition 3. *Let $\mathcal{S}$ be the set of* ***segmented*** *money markets equilibria, defined as* $\{\mathcal{M}(x) \in \mathcal{S} \mid r^i(x) - r^b(x) > \lambda\theta^b\psi(x) \text{ and } r^i(x) - r^p(x) < \lambda\theta^p\psi(x) + \chi\}$.

Definition 4. *Let $\mathcal{A}$ be the set of* ***arbitraged*** *money markets equilibria, defined as* $\{\mathcal{M}(x) \in \mathcal{A} \mid r^i(x) - r^b(x) > \lambda\theta^b\psi(x) \text{ and } w^p(x) < 0\}$.

Definition 5. *Let $\mathcal{E}$ be the set of (traditional bank)* ***reserves satiated*** *equilibria, defined as* $\{\mathcal{M}(x) \in \mathcal{E} \mid \psi(x) = 0\}$.

We define as *integrated* the set of equilibria in which traditional banks are marginal in T-bills; as *segmented* the ones in which traditional banks are not marginal in both T-bills and repos; and as *arbitraged* the ones in which traditional banks are marginal in repos but not in T-bills. We also define the set of equilibria in which traditional banks are *reserves satiated*. All proofs of lemmas and propositions are relegated to Appendix C.

3.1 No Balance Sheet Cost Benchmark

This section describes an economy in which there is no balance sheet cost as a reference point for the analysis. This would correspond to a pre-GFC regime when non-risk-weighted leverage

ratios were not binding. In that case, we show that markets are never segmented so that marginal liquidity risk is equalized across banking sectors.

Proposition 1. *In an economy with no balance sheet cost ($\chi = 0$), money markets are always integrated: $\forall x$ $\mathcal{M}(x) \in \mathcal{I}$, and traditional and shadow banks have the same exposure to liquidity risk per unit of wealth:*

$$\psi(x) = \overline{\psi}(x).$$

Proposition 1 has an intuitive interpretation: risk-averse agents exploit the benefits of risk-sharing by trading liquid assets in order to equalize their marginal utility of holding these liquid assets. Without a balance sheet cost, both agents are always marginal in the market for all liquid assets. This equilibrium condition further implies that banks of all types have the same exposure to liquidity risk and that the joint allocation of repo and T-bills between banks and shadow banks is indeterminate.

Corollary 1. *In an economy with no balance sheet cost ($\chi = 0$), the portfolio weights $w_t^b, w_t^p, \overline{w}_t^b$, and $\overline{w}_t^p$ are jointly indeterminate.*

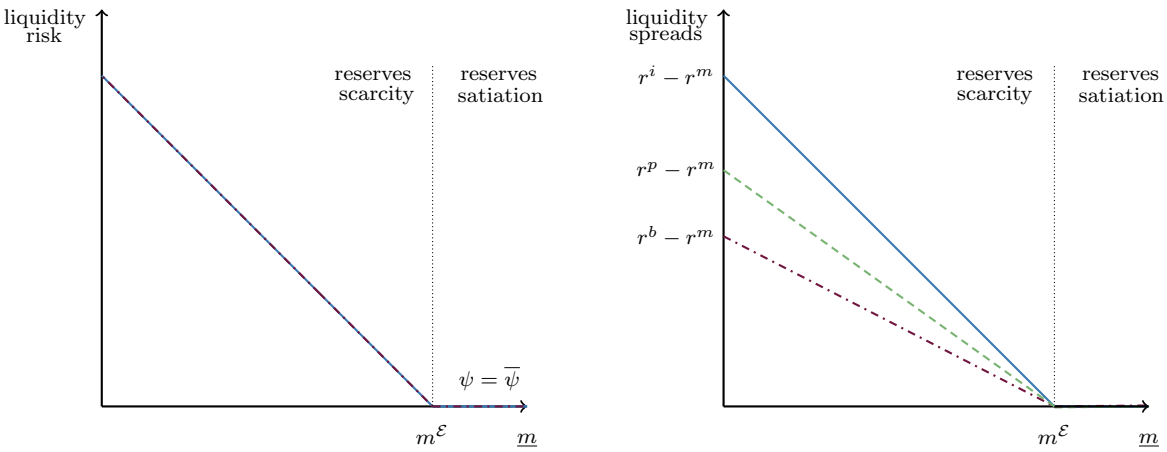
According to Corollary 1, it is the overall exposure to liquidity risk that matters for banks. Since T-bills and repos are perfect substitutes, the same distribution of liquidity risk between the two sectors can be achieved through various combinations of these two assets. Shadow banks could, for example, sell half of their T-bills to traditional banks and receive a similar amount of effective liquidity in the form of repos from traditional banks without impacting any other equilibrium variables. In contrast, traditional banks' portfolio weight on reserves is still determined by the market-clearing condition for reserves, since traditional banks are the only agent that can hold reserves.

To understand the mechanics of the model, consider an expansionary open market operation comparative statics on $m(x)$ in Proposition 2: The central bank increases the supply of reserves by purchasing T-bills. Since reserves are more liquid than T-bills, the net impact of this operation is to increase the effective supply of liquidity in the economy.

Proposition 2. *Consider a monetary policy rule m(x) in an economy with no balance sheet cost ($\chi = 0$) and traditional banks are not satiated with reserves, $\mathcal{M}(x) \in \mathcal{E}^c$. For any given x, an equilibrium with more reserves $m^{\star\star}(x) > m^{\star}(x)$ implies:*

- *less liquidity risk: $\overline{\psi}(x; m^{\star}) = \psi(x; m^{\star}) > \psi(x; m^{\star\star}) = \overline{\psi}(x; m^{\star\star})$,*
- *a lower premium on reserves: $r^i(x; m^{\star}) - r^m(x; m^{\star}) > r^i(x; m^{\star\star}) - r^m(x; m^{\star\star})$,*
- *a lower premium on T-bills: $r^i(x; m^{\star}) - r^b(x; m^{\star}) > r^i(x; m^{\star\star}) - r^b(x; m^{\star\star})$,*
- *a lower premium on repos: $r^i(x; m^{\star}) - r^p(x; m^{\star}) > r^i(x; m^{\star\star}) - r^p(x; m^{\star\star})$.*

Proposition 2 implies that an increase in the supply of reserves uniformly reduces liquidity risk and liquidity premia. This effect appears clearly when taking the partial derivative of the



(a) Liquidity Risk **(b)** Liquidity Spreads

Figure 3: Reserves Supply in Integrated Markets

function for liquidity risk with respect to the policy rule for reserves:

$$\frac{\partial \psi(x;m)}{\partial m} = \frac{\partial \overline{\psi}(x;m)}{\partial m} = \underbrace{\frac{\lambda \theta^b}{n+\overline{n}}}_{\substack{\text{fewer}\\ \text{T-bills}}} - \underbrace{\frac{\lambda \theta^m}{n+\overline{n}}}_{\substack{\text{more}\\ \text{reserves}}} < 0.$$

The effect of a change in the supply of reserves m to the liquidity of traditional banks has two parts. The first term is the higher holdings of reserves by traditional banks. The second term is the reduction in the supply of T-bills available to banks. The net effect is to decrease liquidity risk in both sectors since the liquidity provided by reserves is superior to the liquidity provided by T-bills. As a consequence of Proposition 1, increasing reserves also implies a rebalancing of liquidity portfolios across the two banking sectors. Since traditional banks hold more reserves, they sell T-bills or issue more repos to shadow banks so that liquidity risk is perfectly shared.

Figures 3a and 3b illustrate this result. Liquidity risk in the two banking sectors is monotonically decreasing in the quantity of reserves up to the threshold $m^{\mathcal{E}}$, defined as the point at which the liquidity risk of traditional banks reaches zero. From this point onward, liquidity risk does not depend on the supply of reserves. Figure 3b displays liquidity spreads between the rate on the illiquid asset, repos, and T-bills and interest on reserves as decreasing functions of the supply of reserves. Since the marginal value of holding liquid assets is proportional to exposure to liquidity risk, liquidity premia on reserves and T-bills must drop as a reaction to a larger supply of reserves.

3.2 Balance Sheet Cost and Segmented Money Markets

In this section, we analyze how monetary policy affects the economy in equilibrium with a positive balance sheet cost $\chi > 0$. In this setting, it may be uneconomical for traditional banks to intermediate the liquidity derived from holding additional reserves to shadow banks. In this

case, the two banking sectors face different exposure to liquidity risk. This divergence leads to equilibria with unrealized gains from risk-sharing, as shadow banks have larger marginal benefits of holding liquid assets than traditional banks. We first establish that in the non-segmented portion of the state space, a positive balance sheet cost implies a zero repo creation from traditional banks.

Lemma 1. *In an equilibrium in which money markets are integrated, $\mathcal{M}(x) \in \mathcal{I}$, and which features a strictly positive balance sheet cost $\chi > 0$, traditional banks' equilibrium repo allocation is nil: $w^p(x) = 0$.*

This lemma is intuitive: A traditional bank portfolio allocation that is simultaneously long in T-bills and short in repos is always dominated by a less balance sheet-intensive allocation with only T-bills, no repo, and the same liquidity risk exposure. We characterize the set of segmented equilibria in Proposition 3.

Proposition 3. *In an equilibrium in which money markets are segmented, $\mathcal{M}(x) \in \mathcal{S}$, repo holdings are nil, $w^p = \overline{w}^p = 0$, and shadow banks have a larger exposure to liquidity risk per unit of wealth than traditional banks not accounting for balance sheet cost and lower once accounting for balance sheet cost:*

$$\psi(x) < \overline{\psi}(x) < \psi(x) + \chi/(\lambda\theta^p).$$

Markets for liquid assets are segmented when shadow banks hold all of the supply of T-bills but the demand for repo from shadow banks is not large enough to overcome the balance sheet cost. In this case, the marginal liquidity risk for shadow banks determines the liquidity premium on T-bills and repos, while the marginal liquidity risk for traditional banks determines the liquidity premium on reserves. The pricing factor of the two assets is therefore disconnected, such that the supply of reserves only matters for the liquidity premium on reserves, and the supply of T-bills only matters for the liquidity premium on T-bills. The following proposition characterizes how open market operations affect liquidity risk and liquidity premia when money markets are segmented.

Proposition 4. *Consider a set of monetary policy rules $m(x)$ for a subset of equilibria in which money markets are segmented and traditional banks are not liquidity satiated, $\mathcal{M}(x) \in \mathcal{S} \cap \mathcal{E}^c$. For any given x, an equilibrium with more reserves $m^{\star\star} > m^{\star}$ implies:*

- *less liquidity risk for traditional banks: $\psi(x; m^{\star}) > \psi(x; m^{\star\star})$,*
- *more liquidity risk for shadow banks: $\overline{\psi}(x; m^{\star}) < \overline{\psi}(x; m^{\star\star})$,*
- *a lower premium on reserves: $r^i(x; m^{\star}) - r^m(x; m^{\star}) > r^i(x; m^{\star\star}) - r^m(x; m^{\star\star})$,*
- *a larger premium for T-bills: $r^i(x; m^{\star}) - r^b(x; m^{\star}) < r^i(x; m^{\star\star}) - r^b(x; m^{\star\star})$,*
- *a larger premium for repos: $r^i(x; m^{\star}) - r^p(x; m^{\star}) < r^i(x; m^{\star\star}) - r^p(x; m^{\star\star})$.*

Taking the partial derivative of liquidity risk exposure with respect to the supply of reserves highlights the disconnection in the pricing kernel of the two liquid assets:

$$\frac{\partial \psi(x;m)}{\partial m} = -\frac{\lambda \theta^m}{n} < 0, \qquad \frac{\partial \overline{\psi}(x;m)}{\partial m} = \frac{\lambda \theta^b}{\overline{n}} > 0.$$

In an expansionary open market operation, liquidity risk decreases for traditional banks but increases for shadow banks. From the previous proposition, we can infer that the quantity of reserves may shift an equilibrium from the integrated to the segmented region, and then eventually to the arbitraged region. First, as a reaction to an increase in reserves, traditional banks sell T-bills to shadow banks up to fully depleting their portfolios and hitting their nonnegativity constraint. At this point, $m^{\mathcal{S}}$, the non-negativity T-bill holding constraint becomes binding while repo borrowing is not yet profitable. Then, as the central bank further increases reserves and removes T-bills through open market operations, the demand for repo from shadow banks reaches a second threshold at which repo issuance becomes profitable $m^{\mathcal{A}}$. We formally define these thresholds and characterize a set of equilibria for which the condition $0 < m^{\mathcal{S}} < m^{\mathcal{E}} < m^{\mathcal{A}}$ holds; that is, traditional banks are satiated only when markets are segmented, which takes place before repo arbitrage becomes profitable for banks.

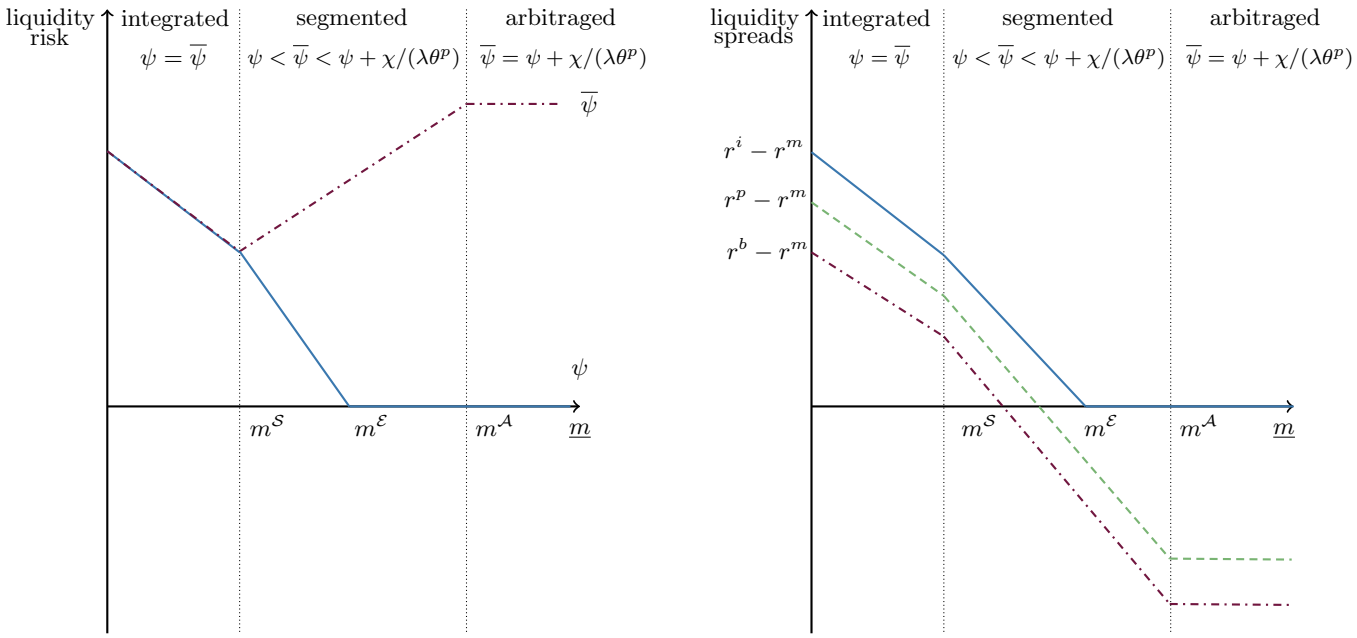
Definition 6. *Define the segmentation threshold $m^{\mathcal{S}}$, arbitrage threshold $m^{\mathcal{A}}$, and satiation threshold $m^{\mathcal{E}}$ as the minimum reserves quantity within each respective sets: $m^i := \min\{m(x) \in i\}$ for $i = \{\mathcal{S}, \mathcal{E}, \mathcal{A}\}$.*

Proposition 5. *Consider a set of monetary policy rules $m(x)$ for a subset of equilibria for which $m^{\mathcal{S}} < m^{\mathcal{E}} < m^{\mathcal{A}}$. For any given x, liquidity risk can be expressed as a function of monetary policy as*

$$\psi(x;m) = \begin{cases} \lambda\left(\sigma^d \frac{1-n-\overline{n}+\tau^h}{n+\overline{n}} - \theta^b \frac{b-m}{n+\overline{n}} - \theta^m \frac{m}{n+\overline{n}}\right), & \text{if } m < m^{\mathcal{S}} \\ \lambda\left(\sigma^d \frac{\gamma}{n}(1-n-\overline{n}+\tau^h) - \theta^m \frac{m}{n}\right), & \text{if } m^{\mathcal{S}} < m < m^{\mathcal{E}} \\ 0, & \text{if } m > m^{\mathcal{E}} \end{cases}$$

$$\overline{\psi}(x;m) = \begin{cases} \lambda\left(\sigma^d \frac{1-n-\overline{n}+\tau^h}{n+\overline{n}} - \theta^b \frac{b-m}{n+\overline{n}} - \theta^m \frac{m}{n+\overline{n}}\right), & \text{if } m < m^{\mathcal{S}} \\ \lambda\left(\sigma^d \frac{1-\gamma}{\overline{n}}(1-n-\overline{n}+\tau^h) - \theta^b \frac{b-m}{\overline{n}}\right), & \text{if } m^{\mathcal{S}} < m < m^{\mathcal{A}} \\ \chi/(\lambda\theta^p), & \text{if } m > m^{\mathcal{A}}. \end{cases}$$

According to Proposition 5, liquidity risk exposures are continuous functions with kinks in the supply of liquid assets. Figure 4 provides a graphical representation of these (comparative-statics) relations between the supply of reserves, liquidity risk, and liquidity spreads. When the supply of reserves is below the threshold $m^{\mathcal{S}}$, the equilibrium is similar to the previous section without balance sheet cost: An increase in reserves supply yields a decrease in liquidity risk for both sectors. Starting from $m^{\mathcal{S}}$, further injections of reserves improve the liquidity risk of the traditional banks at the expense of shadow banks. Since T-bills are the only asset that



(a) Liquidity Risk **(b)** Liquidity Spreads

Figure 4: Reserves Supply in Segmented Markets

shadow banks can own to mitigate liquidity risk, an open market operation that removes T-bills deteriorates their liquidity position. Visually, this effect is seen with the dotted red line moving away from the continuous blue line. From the satiation threshold $m^{\mathcal{E}}$ onward, the liquidity risk of traditional banks reaches and remains at zero. Proceeding to further open market operations after this point increases the liquidity risk of shadow banks without any benefits for traditional banks up to a point $m^{\mathcal{A}}$. Beyond that point, repo spreads are large enough to compensate banks' balance sheet cost χ. These portions of the graph are empirically relevant to analyzing the post-GFC money markets within the new "ample reserves" monetary implementation regime. We discuss monetary policy regimes in the next section.

3.3 Monetary Policy and T-Bill Supply

In this section, we investigate the behavior of rates on liquid assets for two inflation-targeting implementation regimes by the central bank. We first show that when the central bank has the authority to pay interest on reserves, there is a degree of freedom in its target function. This feature implies that multiple monetary policy implementation regimes are feasible. We discuss how the segmentation interacts with these regimes. When money markets are segmented, the central bank does not have to offset changes in the supply of T-bills from the Treasury to stabilize inflation. This result entails that the liquidity premium on T-bills and repos reacts to the supply of T-bills if and only if money markets are segmented.

Monetary Policy Implementation As shown in previous sections, the central bank is able to manage the liquidity premium on reserves by controlling the net supply of liquid assets

through its balance sheet. However, the central bank has a second tool in the model because it also decides on the nominal interest to pay on reserves i^m. This excess in the number of policy tools gives rise to a degree of freedom in the objective function of a central bank that is only interested in controlling inflation.

Lemma 2. *Given x, for any monetary policy rule couple $\{m(x), i^m(x)\}$ able to implement a given inflation target $\pi^\star$: $\pi(x; i^m(x), m(x)) = \pi^\star$, there exists a linear combination of $m(x)$ and $i^m(x)$ that implements the given target $\pi^\star$.*

This result becomes intuitive when substituting for the liquidity premium on reserves in the definition of the nominal rate on reserves $i^m(x) = r^m(x) + \pi(x)$. Doing so yields the following Fischer equation:

$$\pi(x) = \underbrace{i^m(x) + \theta^m \lambda \psi(x)}_{\text{nominal illiquid rate}} - r^i(x). \tag{25}$$

As both the nominal interest on reserves $i^m(x)$ and the liquidity premium $\psi(x)$ are under the control of the central bank, there is one degree of freedom in implementing a given inflation target. This theoretical insight has a straightforward institutional counterpart. In the post-crisis period, the Fed started paying interest on reserves to raise rates and target inflation without having to remove all the reserves that had been created as a byproduct of its large-scale asset purchase program. In the model, this case corresponds to an equilibrium in which banks are fully liquidity-satiated, so liquidity risk is zero for traditional banks ($\psi(x) = 0$).[12] In contrast, in the pre-crisis period, interest on reserves was fixed at a nominal zero rate, as the Fed had no authority to pay any interest on reserves. In this case, the central bank is limited to using market operations to change the liquidity premium on reserves and reach its inflation target.

In Appendix D, we show that in a New Keynesian extension of the model with labor and nominal friction to adjust prices, the same underdetermination is present and that all results presented below are robust to this setting. Intuitively, the New Keynesian block determines how the nominal illiquid rate impacts the output gap and inflation, while the rest of the model focuses on the liquidity management problem of banks and the subsequent liquidity premia of money market instruments.

Central Bank Reaction Function In this section, we consider the interest rate reaction to a T-bill supply shock when monetary policy is endogenous. We show that under both monetary policy implementation regimes, the central bank needs to completely offset the T-bill supply shock when money markets are integrated but not when money markets are segmented. The intuition is that an increase in T-bills boosts the net supply of liquid assets and, in turn, decreases liquidity risk. If the central bank does not react, the nominal rate on illiquid assets must decrease to reflect the lower marginal value of liquid assets. Hence, to keep inflation on target, the central bank has to sterilize the surge in liquidity created by the Treasury by

[12] Woodford (2003) discusses at length why this implementation regime in which there is no role for the supply of money is a satisfactory approximation in Neo-Keynesian models that are not concerned with liquidity.

offsetting open market operations. When money markets are segmented, the T-bill supply shock only affects the liquidity premium on T-bills but not on reserves. Hence, the central bank keeps inflation on target without any intervention.

Proposition 6. *Consider an equilibrium in which money markets are integrated and traditional banks are not liquidity satiated, $\mathcal{M}(x) \in \mathcal{I} \cap \mathcal{E}^c$. Any policy rule $\{m(x), i^m(x)\}$ such that interest paid on reserves is constant $i^m(x) = i^m$ and implementing a constant inflation target π^* fully neutralizes the effect of a change in T-bills on equilibria: $\mathcal{M}(n, \overline{n}, b^\star) = \mathcal{M}(n, \overline{n}, b^{\star\star})$ for any $b^\star$ and $b^{\star\star}$ such that $\mathcal{M} \in \mathcal{I} \cap \mathcal{E}^c$.*

This proposition can be understood by noting that when money markets are integrated, equation (25) can be written as

$$\pi(x) = i^m + \theta^m \underbrace{\lambda^2 \left(\sigma^d \frac{1 - n - \overline{n} + \tau^h}{n + \overline{n}} - \theta^b \frac{b - m}{n + \overline{n}} - \theta^m \frac{m}{n + \overline{n}} \right)}_{\lambda\psi(x)} - r^i(x). \tag{26}$$

When the nominal interest on reserves i^m is held constant, the liquidity risk of traditional banks $\psi(x)$ will drop when b increases. If the central bank wants to keep inflation on target, it needs to adjust the supply of reserves m to prevent $\psi(x)$ from falling. More precisely, to keep liquidity risk $\psi(x)$ constant, the central bank follows the reaction function

$$dm_t = -\frac{\theta^b}{\theta^m - \theta^b} db_t.$$

This reaction function implies that the central bank will decrease the amount of reserves available to banks to offset any exogenous increase in T-bills. As a byproduct of the withdrawal of reserves, the central bank increases the supply of T-bills available to banks. As T-bills are less liquid than reserves, the net effect of these operations is to decrease the amount of aggregate liquidity and return liquidity risk to its initial position. Overall, this reaction from the central bank ensures that the T-bill supply does not affect any equilibrium outcome. A similar analysis is applied to the case in which banks are fully satiated with reserves in Proposition 7.

Proposition 7. *Consider an equilibrium in which money markets are integrated and traditional banks are liquidity satiated, $\mathcal{M}(x) \in \mathcal{I} \cap \mathcal{E}$. The supply of T-bills does not affect equilibria: $\mathcal{M}(n, \overline{n}, b^\star) = \mathcal{M}(n, \overline{n}, b^{\star\star})$ for any $b^\star$ and $b^{\star\star}$ such that $\mathcal{M}(x) \in \mathcal{I} \cap \mathcal{E}$.*

In an integrated equilibrium, when monetary policy has reached a floor and liquidity risk is zero in both banking sectors ($\psi_t = \overline{\psi}_t = 0$), changes in the supply of T-bills are inconsequential such that the central bank does not have to proceed to offsetting open market operations.

Last, we consider the case of segmented markets and find the liquidity premium on T-bills to be disconnected from the liquidity premium on reserves in that regime. Since the central bank implements its policy target through this liquidity premium on reserves, changes in the supply of T-bills do not have to be sterilized. This result, formalized in Proposition 8, is key

to explaining the effect of T-bill supply on repo spreads. The liquidity premium on T-bills and repos is correlated with the T-bill supply when T-bill rates are below the interest rate on reserves, which our model interprets as indicating that money markets are segmented.

Proposition 8. *When money markets are segmented, (i) the supply of T-bills does not affect the liquidity premium on reserves:* $\lambda\theta^m\psi(n,\overline{n},b^\star) = \lambda\theta^m\psi(n,\overline{n},b^{\star\star})$ *for any* $b^\star$ *and* $b^{\star\star}$ *such that* $\mathcal{M}(x) \in \mathcal{S}$*. Hence, (ii) the central bank keeps inflation on target with a policy rule* $\{m(x), i^m(x)\}$ *that does not react to the supply of T-bills so that the supply of T-bills may affect equilibria:* $\mathcal{M}(n,\overline{n},b^\star) \neq \mathcal{M}(n,\overline{n},b^{\star\star})$ *for some* $b^\star$ *and* $b^{\star\star}$ *such that* $\mathcal{M}(x) \in \mathcal{S}$*.*

3.4 Ample Reserves Regime

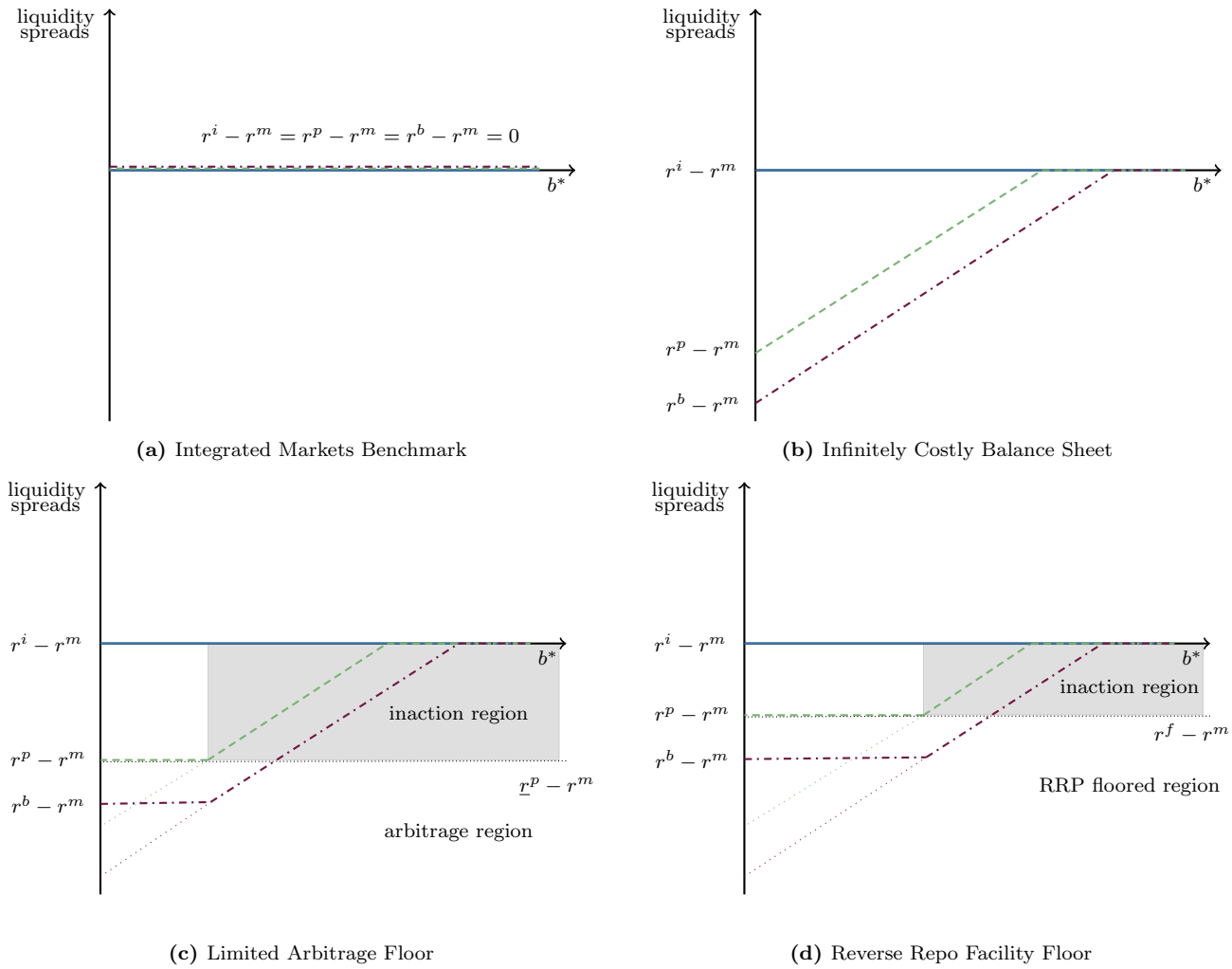
In this section, we apply the model to the empirically relevant context of post-GFC money markets, characterized by traditional banks' liquidity satiation as a side product of large-scale asset purchase programs initiated as a reaction to the GFC and intermittent T-bill shortages. We show how a reverse repo facility—as introduced by the Fed in 2014—can help alleviate obstacles to monetary policy transmission and examine further how various liquidity spreads are affected by changes in T-bill supply.

Reverse Repo Facility Until this point, we have assumed that the central bank was inattentive to liquidity premia on non-reserve money market assets because it does not directly affect its inflation target. The evolution of the post-crisis monetary policy framework of the Fed suggests that this assumption is incomplete. Given the fading relevance of the fed funds market relative to the repo market,[13] the Fed opened a Reverse Repo Facility with a policy rate determined by the FOMC in which shadow banking institutions can deposit funds against eligible collateral. This facility was originally limited in quantities in an experimental phase, but these limits were mostly removed in 2014.[14] This policy translates, in the model, into the central bank standing ready to supply the repo asset elastically to shadow banks at a fixed policy rate, thus effectively borrowing in repo markets. These operations put an effective lower bound on liquid asset rates, as quantities dynamically adjust. We capture this feature by adding the following market-clearing condition for the repo asset in Definition 1:

$$\text{(c) repo:} \quad w_t^p n_t + \overline{w}_t^p \overline{n}_t = f_t.$$

[13] As a consequence of the large amount of excess reserves created by QE, the fed funds volume significantly faded to mostly consist of trades between government-sponsored entities, which do not receive interest on their reserves, and US subsidiaries of foreign banks, with fading relevance for monetary policy transmission.

[14] Although non-binding constraints formally remained for individual users, these were effectively lifted when needed. According to the Fed's website, the overnight reverse repo facility "operate[s] similarly to how the Federal Reserve's payment of interest on excess reserves works for depository institutions. Absent other constraints, any counterparty that is eligible to participate in the ON RRP facility should generally be unwilling to invest funds overnight with another counterparty at a rate below the facility rate. The effectiveness of the facility will depend on a range of factors, including whether a sufficiently broad set of counterparties has access to the facility, the costs associated with regulatory and balance sheet constraints, and the level of competition in the money markets." https://www.newyorkfed.org/markets/rrp_faq_140113.html, retrieved on July 30, 2019.



(a) Integrated Markets Benchmark

(b) Infinitely Costly Balance Sheet

(c) Limited Arbitrage Floor

(d) Reverse Repo Facility Floor

Figure 5: Extended Model Dynamics under Reserves Satiation

The net position of the aggregate banking sector in the repo facility is equal to the quantity of repos supplied by the central bank at the facility f_t. Also, to account for the fact that between 2008 and 2019, the Fed sold all its T-bills and increased the quantity of reserves in the system by buying longer-term assets $\underline{i}_t$, we update the balance sheet identity of the central bank in equation (4) to:

$$\underline{b}_t + \underline{i}_t = m_t + f_t. \tag{27}$$

T-bill Supply and Yields Figure 5 provides a graphical description of how the supply of T-bills affects yields and liquidity premia in four polar cases when traditional banks are liquidity satiated; that is, $m(x)$ is large enough so that $\psi(x) = 0$.

In Figure 5a, there is no balance sheet cost, so arbitrage is costless, and traditional banks can always profitably intermediate liquidity to shadow banks. In this case, there is no market seg-

mentation, and the liquidity satiation of traditional banks translates into the liquidity satiation of shadow banks.

In Figure 5b, the cost of balance sheet space is infinitely high ($\chi \to \infty$), and issuing repos is always too costly for banks. Hence, the only way for traditional banks to trade liquidity is to sell T-bills to shadow banks. While traditional banks are always liquidity-satiated, shadow banks may not be when the supply of T-bills is scarce. The liquidity premium on reserves is zero and independent of the supply of T-bills, while the liquidity premia on repos and T-bills are decreasing functions of the supply of T-bills. Because the illiquid rate r_t^i is fixed to the interest on reserves r_t^m, the rates on T-bills r_t^b and repos r_t^p must adjust upward for the liquidity premium on these assets to go down when the supply of T-bills increases.

In Figure 5c, the balance sheet cost is positive and finite ($0 < \chi < \infty$), which generates an inaction region in which money markets are segmented as it is not profitable for traditional banks to issue repos to shadow banks. Inside the inaction region, the liquidity premium on repos is not large enough to compensate for the balance sheet cost. As described above, once the rate on repos has reached a threshold $\underline{r}_t^p$, traditional banks profitably create repos elastically, thereby preventing the repo rate from falling below this point. The existence of profitable arbitrage creates a floor not only for the repo rate at the threshold $\underline{r}_t^p$ but also for the T-bill rate, as these are perfect substitutes for shadow banks. The supply of liquid assets produced by traditional banks, therefore, matters for the pricing of all liquid assets that can be held by shadow banks.

Figure 5d considers the case in which the central bank introduces its own floor by standing ready to borrow in repos at a rate r_t^f. When doing so, the institution creates any amount of repos necessary to prevent the rate from falling below this threshold. In the figure, the floor set by the central bank is above the one created by the region of profitable arbitrage for traditional banks $r^f > \underline{r}^p$. The region where the spreads are upward sloping is therefore reduced, compared with Figure 5c. In a fixed-rate full-allotment facility, supply must elastically adjust to demand. Assuming that $\underline{i}_t$ and $\underline{b}_t$ remain constant, the expression for the movement in quantities at the reverse repo facility is given by

$$f_t = \left[\frac{\sigma^d(1-\gamma)(1-n-\overline{n}+\tau^h)}{\theta^p} - \frac{\theta^b}{\theta^p} b_t - \frac{r_t^m - r_t^f}{(\lambda\theta^p)^2} \right]^+ . \tag{28}$$

While the central bank borrows whatever quantity is necessary to prevent the repo rate from dropping below the policy rate, r_t^f, it does not lend at this rate. This asymmetry is captured by the maximum in equation (28). When the market rate moves strictly above the policy floor $r_t^p > r_t^f$, there is no demand at the facility rate, and volumes at the facility drop to zero. Equation (28) has an intuitive interpretation. Repo facility intakes are proportional to the net quantity of liquidity corresponding to the administered rate. The higher this rate r_t^f is relative to the rate paid on reserves r_t^m, the more the Fed has to create repos to prevent rates from falling for a given quantity of T-bills. This adjustment is scaled by a factor $(\lambda\theta^p)^{-2}$, reflecting the price elasticity of repos.

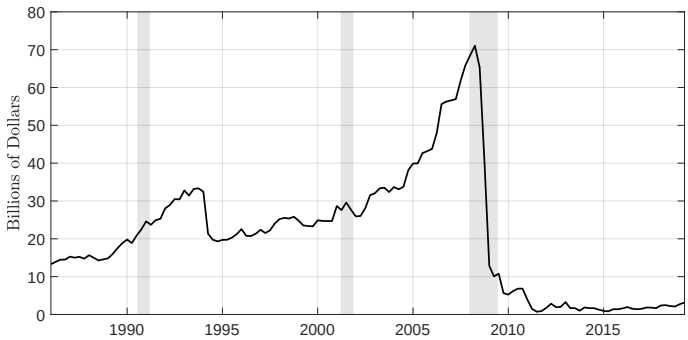


Figure 6: Average Daylight Overdraft Volume at the Fed. The figure displays the evolution of average volumes of intraday overdrafts from the Fed to depository institutions between 1986 and 2021. The series is retrieved from the Payment Systems section of the Federal Reserve Board's website.

4 Quantitative Analysis

In this section, we propose a quantitative application of our model to analyze post-GFC money markets. We first discuss the main challenges in estimating our model and how we address those. Consistently across several specifications, we find that the model accurately predicts movements in money market spreads when rates are above the Fed repo facility floor and volumes at the Fed RRP facility when below. We then use our estimates to investigate counterfactual scenarios for alternative regulation and policy.

4.1 Estimation Strategy

There are four main challenges in estimating our model. First, we do not observe the illiquid rate and, hence, cannot directly determine the prevailing liquidity regime. Second, banks' arbitrage bounds are unobservable and likely heterogeneous. Third, the supply of T-bills is possibly endogenous in money market spreads and other macroeconomic variables. Fourth, some events, such as the money market reform of 2016 and the Covid shock, are likely to act as major confounders through the supply of T-bills and demand for repos. Below, we discuss our measure for excess T-bill supply and how we address these four challenges.

Measuring Excess T-bill Supply We measure the empirical counterpart for the excess supply of T-bills $b_t^\star := b_t - \underline{b}_t$, that is, the supply of T-bills available to shadow banks, as the aggregate nominal outstanding for Treasury securities with maturity below one year at any point in time from TreasuryDirect, subtracting holdings in the Fed SOMA portfolio and a portion of government-only money market funds assets under management (AUM). Thus, $b_t^\star$ = T-bills Outstanding - $\beta\times$ Govt MMF, where β is a parameter to estimate. We make the latter adjustment because government-only money funds are effectively constrained in holding all their assets in either T-bills, repos, or equivalent liquid assets, leaving little room for liquidity

transformation. Doing so is necessary to account for the money market reform of 2014, which led to a significant reallocation within the money market fund sector in favor of government-only money funds. We elaborate on this point below and use this reform as a natural experiment to identify the price elasticity of shadow banks for T-bills.

Evidence of Reserves Satiation The illiquid rate in our model does not have a clear empirical counterpart in our analysis.[15] Hence, we rely on secondary evidence to determine the prevailing liquidity regime. First, we derive information from bank average daylight overdrafts at the Fed.[16] In the micro-foundations of the liquidity transformation problem described (see Online Appendix), any temporary gap between assets and liabilities following a funding shock will require a positive intraday overdraft for as long as banks are not liquidity satiated. Conversely, when banks are liquidity satiated, the quantity of liquid assets is large enough for traditional banks not to need any daily overdraft. Using data from the Fed, Figure 6 shows that overdraft usage has remained particularly low since 2010, which is indicative of traditional banks' liquidity satiation. In addition, we further refer to the work of Afonso, Giannone, La Spada, and Williams (2022), who provide a structural time-varying estimate of the slope of the reserve demand curve from 2010 to 2021 and find that this slope is mostly flat with exceptions from 2010 to 2011 and 2018 to 2019. Motivated by this evidence, we make the assumption that the economy remained in a liquidity satiation regime with the illiquid rate equal to the interest on reserves $r_t^i = r_t^m$ for the entire sample. We stress that our exercise is only valid for periods when this assumption holds. Unsurprisingly, we find below that the model cannot match the time series for repo spreads for most of 2019 when traditional banks were likely not liquidity satiated.

Finding the Effective Floor In our model, the effective floor on the repo rate is the maximum between the floor set by banks' profitable arbitrage and the floor set by the Fed through the reverse repo facility. The reverse repo facility rate is publicly available on the Federal Reserve Board website while the limited arbitrage floor for banks is more challenging to identify as balance sheet costs are not directly observable and likely heterogeneous across banks. To facilitate our quantitative exercise, we make the assumption that it has not been profitable for banks to engage in repo arbitrage since the introduction of the Dodd-Frank Act in 2010. Although this assumption may not hold for all banks during the entire sample period, we believe it is a reasonable working assumption based on the pervasively large arbitrage opportunities since 2010. Following evidence from Du, Tepper, and Verdelhan (2018) and Andersen, Duffie, and Song (2019), we calibrate the risk-free arbitrage bounds of global banks to be around 25

[15] A potential candidate for this illiquid rate would be to use the Z-spread as in Greenwood et al. (2015) and Lenel et al. (2019). The issue with using this measure is that it lacks precision as it is extrapolated from a term structure model of longer maturity and is, therefore, subject to picking up any high-frequency movement on the entire yield curve.

[16] In the US, all traditional banks are part of the real-time gross settlement system (RTGS) Fedwire. Such type of settlement systems, by nature, don't have any room for transaction netting and, hence, require much larger quantities of intraday reserves than are typically available at the end of the day. The Fed meets this demand by creating reserves flexibly during the day through overdrafts.

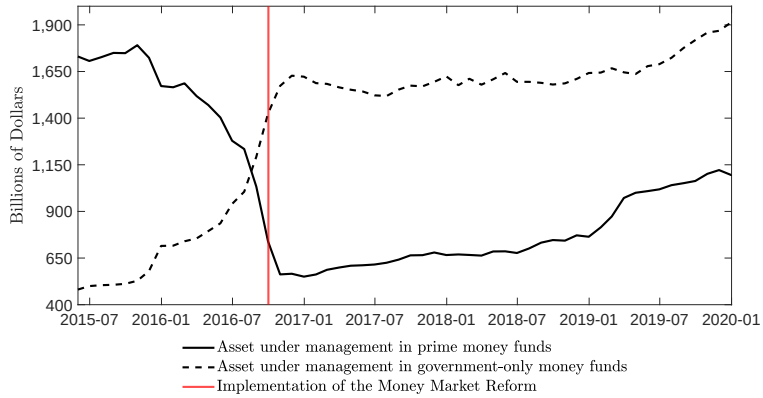


Figure 7: Assets Under Management in Money Funds The figure displays the total value of assets under management in prime (continuous line) and government-only money funds (dotted line) around the implementation of the money market reform in 2016 (vertical line). Source: IMoneyNet

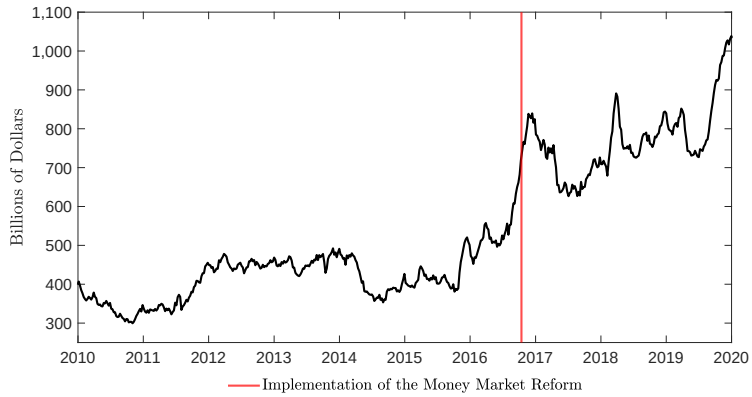


Figure 8: T-bills Held by Money Funds. The figure displays the total amount of T-bills held by money funds around the implementation of the money market reform in 2016. Source: Flow of funds.

basis points. This value exceeds the IOR to triparty repo spread during our entire sample period. Nonetheless, it is possible that for some banks, particularly foreign bank subsidiaries, and before the introduction of Basel III, it was profitable to engage in these arbitrage trades. Our results are robust to this possibility as long as the arbitrage capacity remains limited. However, the presence of marginally unconstrained bank arbitragers with low balance sheet costs could, in theory, bias downward our elasticity estimates.

First Instrument: Tax Calendar In our model, we make the simplifying assumption that the supply of T-bills originating from the Treasury is exogenously determined. However, the maturity structure of Treasury debt, and hence the T-bill supply, is an active decision taken by the Treasury. Specifically, in circumstances where the liquidity premium associated with T-bills is high, the Treasury may be incentivized to increase the volume of short-term T-bill issuance.

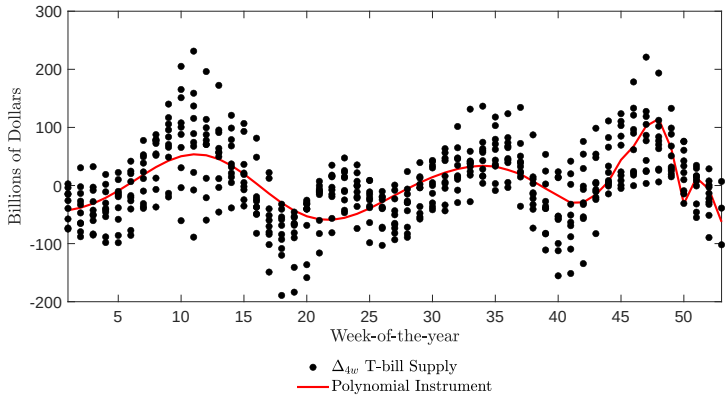


Figure 9: Tax Calendar Polynomial Instrument First Stage Predictions and Observations. The figure displays the prediction of a first-stage regression of an 18th-order polynomial of a week-of-the-year variable on four-week changes in the supply of T-bills outstanding between 2010 and 2020.

Neglecting this consideration runs the risk of inducing a downward bias in our estimations. In addition, macroeconomic variables such as recessionary shocks could further act as confounding variables on T-bill supply and demand for liquid assets. To alleviate these potential concerns, we use an instrumental variable (IV) approach; and, following Greenwood et al. (2015), we exploit the fact that T-bill supply has a strong seasonality pattern due to the plausibly exogenous tax calendar. Specifically, T-bill issuance cycles from low to high as a tax deadline approach as the Treasury deposit account with the Fed is depleted. Applying the Aikake Information Criteria (AIC), we use an 18th-order polynomial of a week-of-the-year variable as an instrument for the supply of Treasury bills. Figure 9 displays the scatter plot and regression prediction for the bivariate first stage of this instrument on T-bills outstanding.

Second Instrument: Money Market Reform Flows Second, as in Cipriani and La Spada (2021) and Anderson, Du, and Schlusche (2021), we exploit the timing of a regulatory change which affected the ability of money market funds to perform their liquidity transformation function. In 2014, the Securities and Exchange Commission (SEC) issued a new set of rules, referred to as the "Money Market Reform" and due for implementation in October 2016. The new rules impose tighter restrictions on portfolio holdings, with an emphasis on liquidity. In particular, an additional requirement from the reform is that prime funds are required to quote a marked-to-market or "floating" value for their shares. This additional rule did not affect government-only funds, which are still authorized to quote a $1 fixed parity. As a consequence, the reform drastically reduced the appeal of prime money funds[17] in favor of government-only funds and triggered an exodus of funds under management from the former to the latter (see Figure 7). Importantly for our analysis, government-only funds are required to hold 80% of their

[17]As argued by Pozsar and Sweeney (2015) and Cipriani and La Spada (2021), prime funds share lost the "moneyness" that made them an attractive investment for corporate treasuries and the cash desks of asset managers.

assets in either T-bills or repos backed by Treasuries, compared with only 30% for prime funds; so that any flow from government-only to prime funds results in an effective reduction in the supply of liquid assets available to remaining shadow banks. As can be seen in Figure 7, around \$1 trillion was reallocated ahead of the reform implementation; Figure 8 shows the associated doubling of money fund T-bill holdings from \$400 billion to around \$800 billion. We exploit this plausibly exogenous shift in the demand for T-bills from the money market fund sector. More precisely, we follow Cipriani and La Spada (2021) in constructing an instrument for government-only flows by regressing government-only inflows on prime outflows at each fund-family level and then aggregating the predicted values for each fund-family each week.

Covid Shock In the reduced-form analysis, we focus on the decade between 2010 and 2020 to avoid having the Covid shock as a confounding factor. In the structural estimation, we extend our sample to 2021 and then control for the Covid crisis with a permanent shock to the liquidity risk of shadow banks σ^d starting on January 1, 2021, which could capture either a potential increase in the risk capacity of shadow banks or an increase in the effective demand for shadow bank deposits.[18]

Money Market Equations Given the above assumptions, we can summarize our three main equations of interest as follows.

$$r_t^m - r_t^p = \begin{cases} \tilde{\theta}^p(\tilde{\sigma}_t^d - \tilde{\theta}^b b_t^\star) & \text{when } f_t \leq \underline{f}, \\ r_t^m - r_t^f & \text{when } f_t > \underline{f}, \end{cases} \tag{29}$$

$$r_t^m - r_t^b = \begin{cases} \tilde{\theta}^b(\tilde{\sigma}_t^d - \tilde{\theta}^b b_t^\star) & \text{when } f_t \leq \underline{f}, \\ r_t^m - r_t^f - (\tilde{\theta}^p - \tilde{\theta}^b)\overline{\psi}_t^\star & \text{when } f_t > \underline{f}, \end{cases} \tag{30}$$

$$f_t = \begin{cases} 0 & \text{when } f_t \leq \underline{f}, \\ (\tilde{\sigma}_t^d - \tilde{\theta}^b b_t^\star)/\tilde{\theta}^p - (r_t^m - r_t^f)/(\tilde{\theta}^p)^2 & \text{when } f_t > \underline{f}. \end{cases} \tag{31}$$

These three equations constitute a system of censored linear regressions with the following definitions:

$$\tilde{\theta}^p \equiv \lambda\theta^p/\sqrt{\overline{n}}, \quad \tilde{\theta}^b \equiv \lambda\theta^b/\sqrt{\overline{n}}, \quad \tilde{\sigma}^d \equiv \lambda\sigma^d(1-\gamma)(1-n-\overline{n}+\tau^h)/\sqrt{\overline{n}}, \quad \overline{\psi}_t^\star \equiv \frac{r_t^m - r_t^p}{\tilde{\theta}^p},$$

and $\tilde{\sigma}_t^d = \tilde{\sigma}^d + \tilde{\sigma}^c \times \mathbb{1}\{t \text{ after January 1st, 2021}\}$ where $\tilde{\sigma}^c$ is the long-term impact of Covid to liquidity risk. As discussed above, when repo rates are below that of the repo facility, all adjustments in liquidity risk are absorbed by the creation of liquid assets at the Fed repo facility. When repo rates are above this policy rate, any increase in the supply of T-bills reduces both T-bill and repo liquidity premia. We consider that the lower bound $\underline{f}$ has been

[18] Using a different timing for this dummy variable would not materially change our quantitative results as long as it is located during the long period of shadow bank liquidity satiation of post-March 2020 to early 2021.

reached when the volume at the repo facility is below $15 billion. Moreover, as our definition of shadow banks would also include offshore institutions participating in money markets, it is challenging to obtain a good proxy for shadow banks' net worth. Hence, we make the working assumption that the net worth share of shadow banks is constant over time and let the empirical observations on spreads inform us whether this simple model is rejected or not. We propose three quantitative exercises. We first estimate these three regressions separately between 2010 and 2020 to provide reduced-form evidence, including additional controls and instrumental variable regressions. We then conduct a structural exercise by estimating these equations with and without the parametric restrictions imposed by the model on an extended sample to 2022. Finally, we estimate an extension of the model that relaxes the assumption that households are fully rate-inelastic to account for some possible substitution between traditional and shadow deposits in relation to movement in the level of interest paid on reserves.

4.2 Reduced-form Evidence

We estimate our main three equations individually within three specifications, all reported in Table 1. In all specifications, we include controls for the VIX and month-end effect as is standard in the money market literature. Moreover, following Nagel (2016) and to account for possible deposit channel effect stemming from the level of interest rates, we also control for the level of interest rates by adding the interest on reserves as a control variable. In the first (level) specification, we use a Tobit regression model to account for the censoring described above and report the parameter estimates in the first three columns. As expected from our theory, the regression finds a positive association between the supply of Treasury bills and both the T-bill-IOR and the TGCR-IOR spread, as well as a negative relation with quantities at the RRP. In terms of magnitudes, an increase in $100 billion in Treasury bills outstanding is associated with an increase of 4 basis points in Tbill-IOR spreads, 4 basis points in TGCR-IOR spreads, and a decrease in RRP volumes of $12 billion. Similarly, we find that an increase in government-only money market funds AUM is negatively associated with the two spreads and positively associated with RRP intake. We also note that IOR has a positive effect on spreads but is only slightly statistically significant on the TGCR-IOR spread. In Appendix G, we display the predictions of these Tobit models against realized observations for T-bill-IOR, TGCR-IOR, and RRP volumes. To alleviate the possible concern that the relationship between the supply of T-bills and money market spreads could be driven by a stochastic trend generating a spurious correlation between the level of these variables, we also run our analysis in a four-week time-differenced specification. Columns 4 to 6 report the result for these regressions. Although the estimate on the T-bill and repo spread regressors are reduced, our estimates remain consistent with the theory and in close range of the Tobit regression estimates. We also note that the IOR is only statistically significant on the T-bill-IOR spread with a negative sign, which is consistent with a repricing effect of 1-month T-bills after a monetary policy shock changing the IOR. The IOR not being statistically significant on repo spreads is indicative of a muted deposit channel effect, which we confirm below.

Table 1: Baseline Regressions. For each triplet group, the table presents results from weekly regressions of 1-month T-bill (T-bill-IOR) and Overnight triparty (TGCR-IOR) yield spreads in excess of IOR and of Quantities at the Reverse Repo Facility (RRP Vol.) on outstanding supply of Treasury bills (T-bill Supply), government-only money market funds assets under management (Govt MMFs), and the interest on reserves (IOR). Columns (1)-(3) show results from a Tobit regression in a level specification. Control variables include: the VIX, a linear trend, and an end-of-month calendar fixed effect. Columns (4)-(6) report OLS estimates where all regression variables are expressed in a 4-week difference. Columns (7)-(9) present results from GMM regressions with all variables expressed in a 4-week difference. The endogenous variables are ΔT-bill Supply and ΔGovt MMFs and are instrumented using week-of-year polynomial dummies up to 18th degree and an instrumental variable based on the Prime-to-Government MMF Reform weekly flow. HAC adjusted standard errors, and GMM weighting matrix for the 4-week difference regressions are estimated using the Newey-West estimator with 12 lags, and Control variables include: the VIX and beginning- and end-of-month calendar fixed effects. The table row Sample denotes the regression estimation sample period split by the volumes at the RRP facility. The Cragg-Donald Wald F-statistic of the GMM procedure is 26.27.

	Level Tobit			4-Week Diff OLS			4-Week Diff GMM		
	(1) T-bill-IOR	(2) TGCR-IOR	(3) RRP Vol.	(4) ΔT-bill-IOR	(5) ΔTGCR-IOR	(6) ΔRRP Vol.	(7) ΔT-bill-IOR	(8) ΔTGCR-IOR	(9) ΔRRP Vol.
T-bill Supply ($bln.)	0.039***	0.037***	−0.122***	0.029**	0.025**	−0.300***	0.020**	0.025***	−0.188***
	(0.002)	(0.002)	(0.018)	(0.011)	(0.008)	(0.063)	(0.007)	(0.006)	(0.054)
Govt MMFs ($bln.)	−0.012***	−0.016***	0.116***	−0.034*	−0.025*	0.281***	−0.036***	−0.020	0.251***
	(0.001)	(0.001)	(0.009)	(0.016)	(0.012)	(0.062)	(0.010)	(0.012)	(0.046)
IOR	0.582	2.437*	10.40	−27.27***	3.781	150.0***	−29.42***	3.177	126.8***
	(1.034)	(1.188)	(6.888)	(5.236)	(3.760)	(37.45)	(4.791)	(2.760)	(27.90)
Intercept	−83.82***	−74.14***	875.8***	0.476	0.120	−8.676*	0.719*	−0.086	−7.024*
	(5.871)	(6.478)	(101.9)	(0.427)	(0.321)	(4.121)	(0.343)	(0.233)	(2.906)
# Obs.	522	522	261	304	304	218	304	304	218
Period	2010-2019	2010-2019	2015-2019	2010-2019	2010-2019	2013-2019	2010-2019	2010-2019	2013-2019
Sample	Full	Full	Full	≤ $15bn.	≤ $15bn.	> $15bn.	≤ $15bn.	≤ $15bn.	> $15bn.
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

*p<0.05, ** p<0.01, *** p<0.001

We report in columns 7 to 9 the results for the four-week time-differenced specification by GMM using the instrumental variable strategy described above. The first stage of this procedure reports a Cragg-Donald Wald F-statistic of 26.27, which satisfies Stock and Yogo's (2005) test for weak IVs. Most estimates remain robust to using the IV strategy in retaining the expected sign and some level of statistical significance. The coefficient for T-bill-IOR on T-bills outstanding is reduced to 0.0197, while the coefficient for T-bill-GCR on the same independent variable remains stable at 0.0255. The coefficient on RRP quantity intake is also reduced to -0.188. The reduction in the T-bill-IOR estimate suggests that confounding effects on the timing of T-bill issuance are more prevalent than strategic considerations from the Treasury. Coefficients on government-only money funds AUM are remarkably stable, despite the estimate on TGCR-IOR losing statistical significance, suggesting that there are little spillovers from money market spreads to money market fund flows.

In Table 3, we also report the same regressions while controlling for the supply of outstanding reserves. Overall, except for the level-estimate on TGCR-IOR spread being reduced, all estimations remain robust to the inclusion of reserves as an additional control.

4.3 Structural Analysis

In this section, we estimate the three equations (29), (30), and (31) both with and without the cross-equation parametric restrictions implied by the model. Naturally, by being more flexible, the unrestricted model outperforms the constrained model in terms of fitting observations. The objective of this exercise is to provide an assessment for the theoretical prediction that the three time series are jointly integrated by the common element $\overline{\psi}$, shadow banks' demand for liquid assets.

Table 2 reports the results of our estimations. An increase in T-bill supply of \$100 billion generates an increase in the repo-IOR and T-bill-IOR spreads of 6 basis points in the constrained model. Once having reached the RRP rate, a further decrease in T-bill supply of \$1 leads to an increase \$1 in RRP volume. All these figures are higher in the constrained than in the reduced-form regressions because, in the former, all intercepts are derived from a unique $\tilde{\sigma}^d$. Moreover, the post-January-2021 dummy allows us to capture the magnitude of the Covid demand shock and indicates a doubling of shadow banks' aggregate liquidity risk. Panel B provides the results of estimating each regression independently, that is, without the cross-equation parametric restrictions implied by the model.

As shown in Figure 10, the constrained model still accurately predicts significant portions of the post-2010 time series for money market spreads and reverse repo facility volumes. Notably, the steady increase in the stock of T-bills observed in 2018 corresponds to a significant increase in repo and T-bill yields and to a decline in reverse repo volumes to zero. Interpreted through the model's lens, this increase in the T-bill supply led to a reduction in the scarcity of liquid assets for shadow banks and the liquidity premia on money market assets decreased. Noticeably, the model is unable to capture the disconnection between T-bill and repo yields that took

	Panel A: Constrained Regressions			Panel B: Unconstrained Regressions		
	TGCR-IOR (bps)	T-bill-IOR (bps)	RRP Vol. ($bln.)	TGCR-IOR (bps)	T-bill-IOR (bps)	RRP Vol. ($bln.)
T-bills ($bln.)	0.062	0.063	−1.030	0.050	0.030	−0.869
	(0.003)	(0.004)	(0.028)	(0.007)	(0.001)	(0.048)
Govt MMFs ($bln.)	−0.022	−0.022	0.346	−0.015	−0.007	0.334
	(0.001)	(0.002)	(0.015)	(0.003)	(0.001)	(0.023)
Intercept	−96.734	−99.621	1,614.645	−80.000	−61.253	1,258.642
	(3.716)	(4.585)	(48.022)	(9.493)	(0.935)	(79.526)
Covid Dummy	−99.924	−105.906	1,667.896	–	–	1,477.549
	(5.671)	(6.582)	(42.246)	–	–	(84.585)
N	670	670	670	670	670	670

Estimated parameters:	$\tilde{\theta}^p = 0.245$	$\tilde{\theta}^b = 0.252$	$\tilde{\sigma}^d = 395.21$	$\tilde{\sigma}^c = 408.24$	$\beta = -0.354$
Standard errors:	(0.007)	(0.007)	(6.485)	(13.522)	(0.008)

Table 2: Structural Estimations. At the bottom of the table, we provide the estimates of the parameters $\tilde{\theta}^p$, $\tilde{\theta}^b$, $\tilde{\sigma}^d$, $\tilde{\sigma}^c$, and β using maximum likelihood estimation. In Panel A, we also report the implied coefficient corresponding to a reduced-form approach: $y_t = \alpha + \beta^{Tbills} \times$ T-bills Outstanding $+ \beta^{Govt} \times$ Govt MMF $+ \beta^{Covid} \times \mathbb{1}\{t \text{after January 1st, 2021}\}$ assuming that we observe the T-Bill-IOR, TGCR-IOR only if RRP quantities are greater than $15 billion. For example, for the TGCR-IOR regression, the coefficients α, β^{Tbills}, β^{Govt}, and β^{Covid} correspond to $-\tilde{\theta}^p\tilde{\sigma}^d$, $\tilde{\theta}^p\tilde{\theta}^b$, $\tilde{\theta}^p\tilde{\theta}^b\beta$, and $-\tilde{\theta}^p\tilde{\sigma}^c$, respectively. In panel B, we estimate the unconstrained regressions independently for comparison. Reported standard errors in the parenthesis are computed using 2,000 bootstrap samples of 670 observations. Appendix E contains more details regarding the maximum likelihood estimation.

place summer of 2019. This episode is particularly puzzling from the point of view of the model in which the two assets are always perfect substitutes. Although not featured in this model, we hypothesize that a shortage of intraday liquid, as was prevalent at the time as shown by Copeland et al. (2022); d'Avernas and Vandeweyer (2021), could have altered this perfect substitutability and made repos more valuable relative to T-bills. Furthermore, as predicted by the model for the period between March 2020 and January 2021, both repo and T-bill yields are trading close to the IOR, while reverse repo volumes are at zero as a consequence of the large T-bill supply generated by the series of fiscal stimulus packages. This trend was then reversed in 2021 after the Treasury started shifting to issuing longer maturity assets and decreased the supply of T-bills to close to historically low levels.

4.4 Shadow Deposit Channel Extension

In this section, we extend our model to account for deposit channel dynamics on banks and shadow banks shown by Drechsler, Savov, and Schnabl (2017) and Xiao (2020). Households now choose between different liquid assets: cash, shadow bank deposits, and traditional bank deposits with positive elasticity, according to a CES aggregator. We present a full characterization of this model and the estimation methodology in Appendix F and discuss the key insights here.

Description The introduction of this shadow deposit channel generates a non-trivial interaction between changes in banks' market power following movement in the level of interest rates, anchored to the interest on reserves and the pricing of money market instruments. As, in this

setting, banks have some market power due to the finite elasticity of deposit demand to their deposit rates, the first order conditions for deposits (19) and (23) become:

$$-\frac{r_t^d}{d_t^h}\frac{\partial d_t^h}{\partial r_t^d} = \frac{r_t^d}{r_t^d - r_t^i + \chi}, \tag{32}$$

$$-\frac{\overline{r}_t^d}{\overline{d}_t^h}\frac{\partial \overline{d}_t^h}{\partial \overline{r}_t^d} = \frac{\overline{r}_t^d}{\overline{r}_t^d - r_t^i + \lambda\sigma^d\overline{\psi}_t + \overline{\chi}}, \tag{33}$$

where d_t^h and $\overline{d}_t^h$ are the households' demand for traditional and shadow bank deposits, respectively. While upholding our assumption that banks are always satiated with ψ_t equal to 0 in equation (32), we also add a marginal cost for shadow banks to raise deposits $\overline{\chi}$ to account for operational costs and allow for finite elasticity when shadow banks are fully satiated.[19] With these modifications, our extended model captures the previously documented dynamics of deposit demand and markups. Generally, the lower the demand elasticity to deposit rates, the higher banks can set their deposit rates. This mechanism implies that, on the one hand, when the level of interest rates rises, cash becomes a more expensive substitute for deposits, and the market power of traditional banks increases (Drechsler, Savov, and Schnabl, 2017). On the other hand, cash is not particularly substitutable with shadow bank deposits, so their market power does not increase following a rise in interest rates (Xiao, 2020). Consequently, traditional banks with high market power widen their spreads, and deposits flow out in the more concentrated deposit market of shadow banks. In turn, this effect results in an increase in the demand for shadow bank deposits and a subsequent increase in the quantity of liquidity risk for shadow banks $\overline{\psi}_t$, which then impacts other money market rates.

Estimation We estimate this extended model with two additional time series to discipline the estimation of the parameters: the average interest rate for traditional and shadow deposits. As shown in Figure 11, the traditional banks' deposit spread closely follows the interest rate of the Fed, while the shadow banks' deposit spread remains close to constant over time. These additional targets impose further constraints on the model's parameters. First, for the traditional banks' deposit spread to follow the interest on reserves closely, the market power of traditional banks must increase almost one-for-one, and therefore, deposits need to be an important fraction of the liquidity basket of households, and conversely for shadow banks. Second, to reach such a high level, the market power of traditional banks needs to be relatively high compared to shadow banks, in line with the findings of Xiao (2020) for the 1988-2013 period.

Analysis Our estimations of the extended model suggest that during the 2010-2022 period, the deposit channel is not a key driver of money market rates. Indeed, the month during which the fed funds rate is the highest in 2019 corresponds to periods during which money market

[19] When repo or treasury spreads are equal to 0 in our sample, it implies that $\overline{\psi}$ equals 0. With $\overline{\chi}$ also equal to 0, this would imply that the elasticity of the demand with respect to the deposit spreads needs to be equal to 1, which is not possible in a model with CES aggregators with substitute assets, as shown in the derivation of that elasticity in Appendix F.

spreads are close to 0. Consequently, the maximum likelihood estimates converge towards parameter estimates that mute the shadow bank deposit channel, and the model's predictions are nearly the same as the predictions of the benchmark model. Similarly, parameter estimates such that the deposit channel drives a part of the Covid spike would not be consistent with the 2016-2020 period. In Figure 11, we provide the model's predictions for the markups and deposit spreads of traditional and shadow banks.[20]

4.5 Counterfactual Exercise

In this section, we use the quantitative model to derive counterfactual estimates for three scenarios: (i) the absence of the money market reform, (ii) the absence of a reverse repo facility, and (iii) the Fed setting up a repo facility for shadow banks at a rate equal to the interest rate on reserves, starting in 2010.

No Money Market Reform To evaluate the effect of the money market reform on the pricing of money market instruments, we derive a model-implied counterfactual time series for our three main variables of interest from the T-bill supply series, assuming that the ratio of government-only AUM to total money market funds AUM had remained at its November 1, 2015, 39% level. The results are depicted by the red lines in Figure 12. Without the money market reform, which drained around $400 billion of T-bills, yields on T-bills and repos would have already reached the interest rate on reserves at the beginning of 2018. Consequently, the volumes of reverse repos from the Fed would have dropped to zero as early as October 2016.

No Reverse Repo Facility We proceed in a similar way to build a counterfactual time series for an economy in which the Fed did not set a floor on repo rates. The results are depicted by the blue lines in Figure 12. As a consequence of the drop in the supply of T-bills in 2016, yields on repo transactions and T-bills drift further below the IOR. Between 2015 and 2018, both repo and T-bill rates move beyond -40 bps below the IOR. However, the most significant drop is found at the end of our sample, when the T-bill supply is reduced from $5 trillion to $3.5 trillion in less than a year. Our model predicts that the drop in repo rates would have been more than 80 bps below the interest on reserves and hence 70 bps below the lower bound of the Fed's target. Note that this result assumes that banks would still not find it profitable to raise equity to increase their balance sheet for such a large spread.

Reverse Repo Facility Paying IOR Last, we also consider a counterfactual scenario in which the Fed had opened a reverse repo facility to shadow banks in 2010 that paid an interest rate equal to the interest on reserves. As we can see from the yellow lines in Figure 12, doing so would have meant much larger intakes than what was observed in our sample. Intuitively, in order for T-bill and repo rates not to fall below the interest rate on reserves, we need shadow

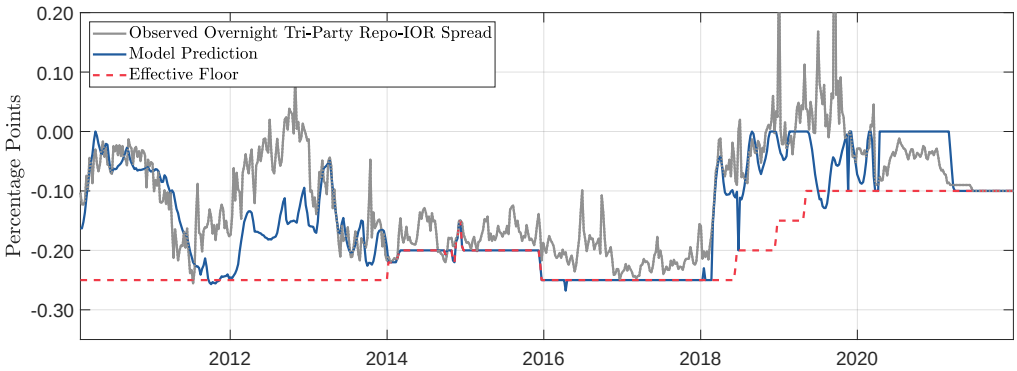
[20] The markups are given by $\left(-\partial \log(d_t^h)/\partial r_t^d\right)^{-1}$ and $\left(-\partial \log(\overline{d}_t^h)/\partial \overline{r}_t^d\right)^{-1}$.

banks not to bear any liquidity risk at any point in time. Recalling equation (28), we can see that the smaller the spread between the repo facility rate r^f and the interest rate on reserve r^m, the more repos need to be created by the Fed. Between 2016 and 2018 when the RRP facility was at 25 bps below IOR, reducing this spread to zero would have translated into an increase in RRP volumes of around \$400 bn. This figure is much lower since 2021, reflecting that the RRP to IOR spread had then been reduced to 10 bps.

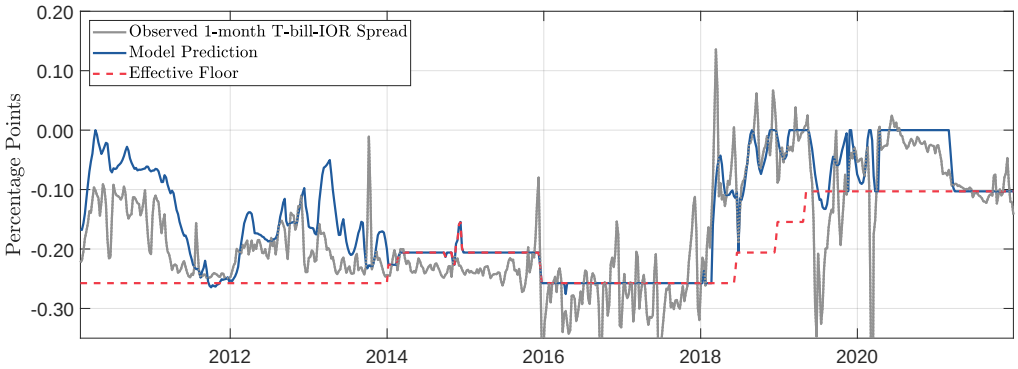
5 Conclusion

The 2008 financial crisis was a reminder that unregulated liquidity creation leads to financial instability. To curb liquidity transformation from banks, more stringent regulations have since been introduced, but this has also impaired their ability to intermediate liquidity to the rest of the financial system. As a result, public institutions had to take on a larger responsibility of providing outside money. This paper focuses on the importance of Treasury-created outside money in a world where liquidity intermediation is costly. We propose a model of local money market segmentation, which demonstrates that intermittent shortages of T-bills are a significant driver for T-bill and repo rates, as well as volumes at the Fed's reverse repo facility. The model illustrates the difficulties central banks face in controlling short-term interest rates beyond strict interbank markets in a post-regulatory-reforms world, and our analysis underpins a strong motive for introducing additional financial instruments allowing access to digital central bank liabilities to shadow banks, such as "Fed bills" or Central Bank Digital Currency. In particular, we estimate that the demand for T-bills or Fed repos combined exceeded \$3.2 trillion in 2021,[21] which would have led to a sharp fall in short-term rates if the RRP facility did not automatically compensate for the shortfall between this demand and the actual T-bill supply.

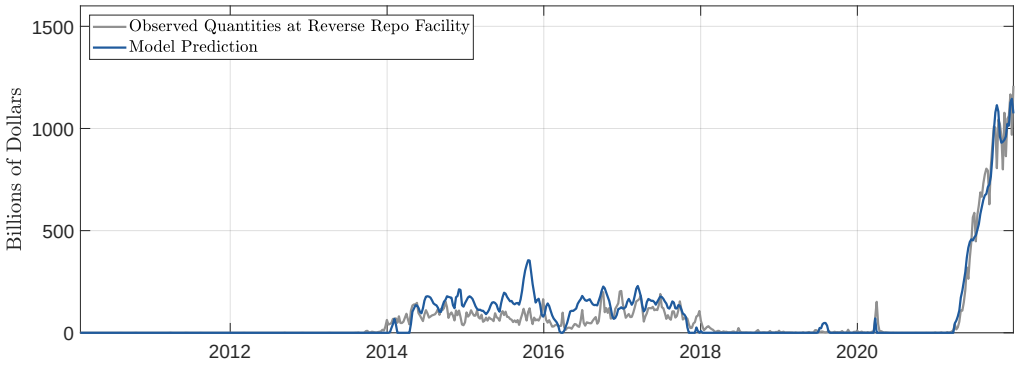
[21] In our structural model estimated in Section 4.3, the total demand for liquid assets after January 1st, 2021, is given by $(\tilde{\sigma}^d + \tilde{\sigma}^c)/\tilde{\theta}^p$.



(a) Model Predictions for the Spread between the Repo Rate and the IOR

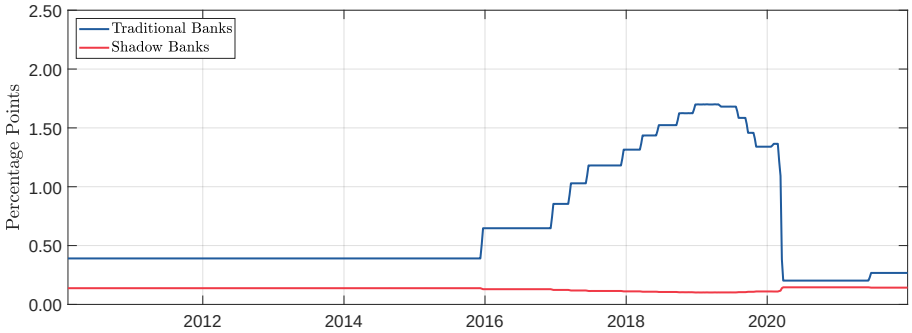


(b) Model Predictions for the Spread between T-bill yield and the IOR

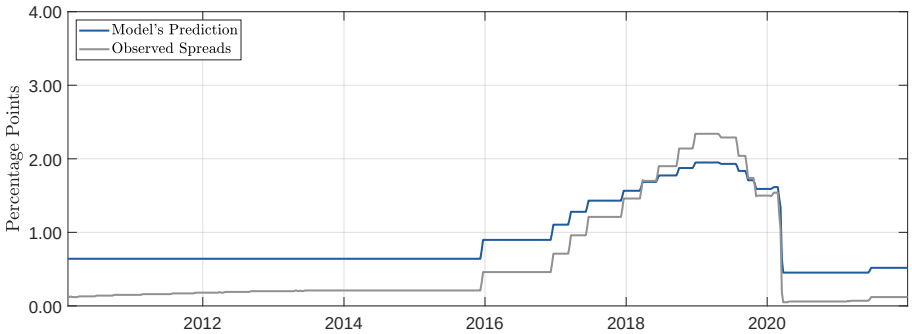


(c) Model Predictions for Volumes at the Reverse Repo Facility

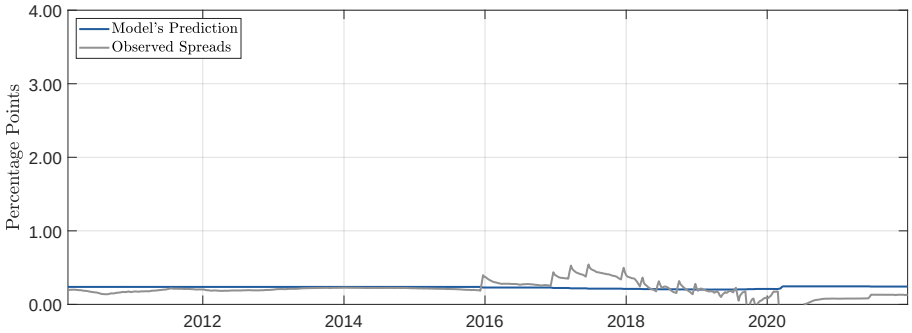
Figure 10: Model Predictions and Weekly Observations. This figure displays the time series for observed weekly data (gray) for T-bill and repo spreads and RRP facility volumes, as well as model-implied predictions.



(a) Traditional and Shadow Banks' Markups

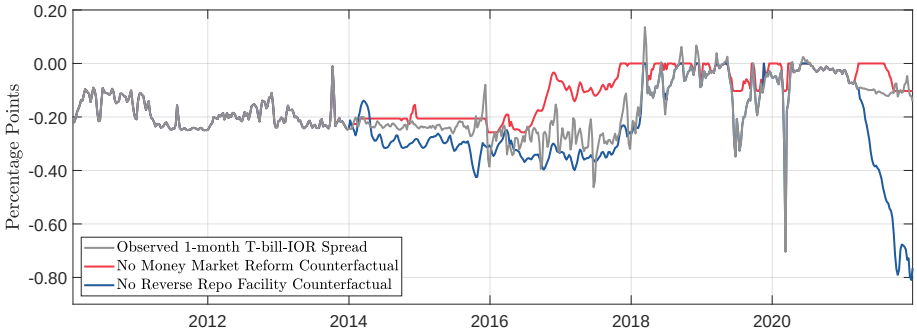


(b) Spread between IOR and Traditional Bank Deposit Rate

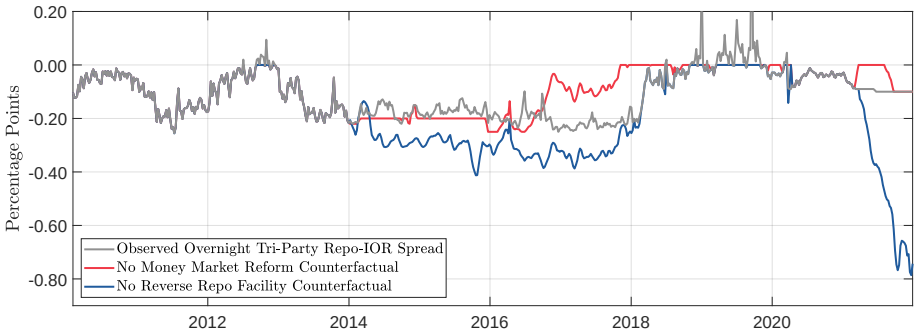


(c) Spread between IOR and Shadow Bank Deposit Rate

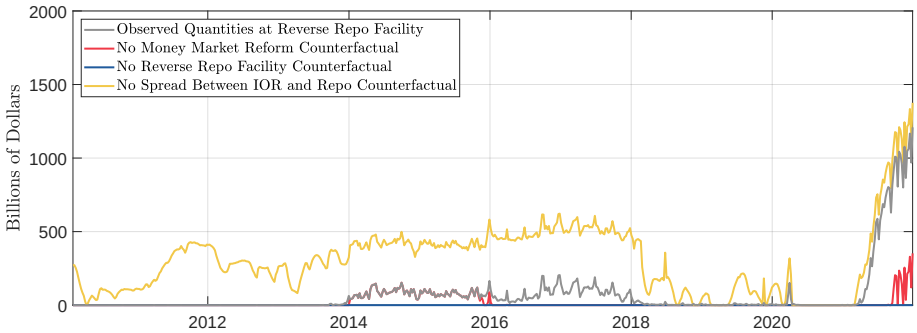
Figure 11: Model Predictions and Weekly Observations for Deposit Channel Extension. This figure displays the time series for observed weekly data (gray) for traditional and shadow banks' deposit spreads and the model's predictions for traditional and shadow banks' markups and deposit spreads.



(a) Spread between the Repo Rate and IOR



(b) Spread between the T-bill yield and the IOR



(c) Volumes at the Reverse Repo and Repo Facility

Figure 12: Counterfactual Analysis. This figure displays the time series for observed weekly data (gray) for T-bill and repo spreads and reverse repo facility volume, as well as model-implied counterfactuals for three scenarios: (i) no money market reform (red), (ii) no reverse repo facility (blue), and (iii) a reverse repo facility with no spread between the IOR and the reverse repo facility rate. Appendix E contains more details regarding the construction of the counterfactual time series.

References

Acharya, Viral and Skeie, David. A model of liquidity hoarding and term premia in inter-bank markets. *Journal of Monetary Economics*, 58(5):436–447, 2011.

Afonso, Gara and Lagos, Ricardo. The over-the-counter theory of the fed funds market: A primer. *Journal of Money, Credit and Banking*, 47(S2):127–154, June 2015.

Afonso, Gara, Giannone, Domenico, La Spada, Gabriele, and Williams, John C. Scarce, abundant, or ample? A time-varying model of the reserve demand curve. Staff Report 1019, Federal Reserve Bank of New York, 2022.

Amstad, Marlene and Martin, Antoine. Monetary policy implementation: Common goals but different practices. *Current Issues in Economics and Finance*, 17(7), November 2011.

Andersen, Leif, Duffie, Darrell, and Song, Yang. Funding value adjustment. *Journal of Finance*, 74(1):145–192, 2019.

Anderson, Alyssa G., Du, Wenxin, and Schlusche, Bernd. Arbitrage capital of global banks. Finance and Economics Discussion Series 2021-032, National Bureau of Economic Research, April 2021.

Andolfatto, David and Williamson, Stephen. Scarcity of safe assets, inflation, and the policy trap. *Journal of Monetary Economics*, 73(C):70–92, 2015.

Arrata, William, Nguyen, Benoît, Rahmouni-Rousseau, Imène, and Vari, Miklos. Eurosystem's asset purchases and money market rates. Technical report, Banque de France, 2017.

Avdjiev, Stefan, Du, Wenxin, Koch, Cathérine, and Shin, Hyun Song. The dollar, bank leverage, and deviations from covered interest parity. *American Economic Review: Insights*, 1(2):193–208, September 2019.

Bansal, R. and Coleman, W.J. A monetary explanation of the equity premium, term premium and risk-free rate puzzles. *Journal of Political Economy*, 104(6):1135–1171, 1996.

Bech, Morten and Monnet, Cyril. A search-based model of the interbank money market and monetary policy implementation. *Journal of Economic Theory*, 164(C):32–67, 2016.

Bech, Morten L. and Klee, Elizabeth. The mechanics of a graceful exit: Interest on reserves and segmentation in the federal funds market. *Journal of Monetary Economics*, 58(5):415–431, 2011. doi: 10.1016/j.jmoneco.2011.05.

Bech, Morten L., Klee, Elizabeth, and Stebunovs, Viktors. Arbitrage, liquidity and exit: The repo and federal funds markets before, during, and emerging from the financial crisis. FEDS Working Paper 2012-21, November 2011.

Bianchi, J. and Bigio, S. Banks, liquidity management and monetary policy. *Econometrica*, 90 (1):391–454, 2022.

Boyarchenko, Nina, Costello, Anna M., and Shachar, Or. The long and short of it: The post-crisis corporate CDS market. *Federal Reserve Bank of New York Economic Policy Review*, 26 (3):1–48, 2020.

Brunnermeier, Markus K. and Pedersen, Lasse Heje. Predatory trading. *Journal of Finance*, 60(4):1825–1863, 2005.

Brunnermeier, Markus K. and Sannikov, Yuliy. A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421, 2014.

Brunnermeier, Markus K. and Sannikov, Yuliy. The i-theory of money. NBER Working Papers 22533, National Bureau of Economic Research, 2016.

Bukhari, Meryam, Cambron, Alyssa, Fleming, Michael J., McCarthy, Jonathan, and Remache, Julie. The SOMA portfolio through time, August 12 2013.

Caballero, Ricardo J. On the macroeconomics of asset shortages. NBER Working Papers 12753, National Bureau of Economic Research, Inc, December 2006.

Cipriani, Marco and La Spada, Gabriele. Investors' appetite for money-like assets: The MMF industry after the 2014 regulatory reform. *Journal of Financial Economics*, 140(1):250–269, 2021.

Copeland, Adam, Duffie, Darrell, and Yang, Yilin. Reserves were not so ample after all. NBER Working Papers 29090, National Bureau of Economic Research, 2022.

d'Avernas, Adrien and Vandeweyer, Quentin. Intraday liquidity and money market dislocations. Working paper, University of Chicago Booth School of Business, 2021.

d'Avernas, Adrien, Vandeweyer, Quentin, and Darracq-Pariès, Matthieu. Unconventional monetary policy and funding liquidity risk. Technical report, 2019.

De Fiore, Fiorella, Hoerova, Marie, and Uhlig, Harald. Money markets, collateral and monetary policy. NBER Working Papers 25319, National Bureau of Economic Research, Inc, November 2018.

Diamond, Douglas and Rajan, Raghuram. Liquidity risk, liquidity creation, and financial fragility: A theory of banking. *Journal of Political Economy*, 109(2):287–327, 2001.

Diamond, Douglas and Rajan, Raghuram. Liquidity shortages and banking crises. *Journal of Finance*, 60(2):615–647, 2005.

Drechsler, Itamar, Savov, Alexi, and Schnabl, Philipp. The deposits channel of monetary policy. *Quarterly Journal of Economics*, 132(4):1819–1876, April 2017.

Drechsler, Itamar, Savov, Alexi, and Schnabl, Philipp. A model of monetary policy and risk premia. *Journal of Finance*, 73(1):317–373, 2018.

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

Duffee, Gregory R. Idiosyncratic variation of treasury bill yields. *Journal of Finance*, 51(2): 527–551, June 1996.

Duffie, Darrell and Krishnamurthy, Arvind. Passthrough efficiency in the Fed's new monetary policy setting. In *Designing Resilient Monetary Policy Frameworks for the Future, A Symposium Sponsored by the Federal Reserve Bank of Kansas City*, pages 21–102, 2016.

Duffie, Darrell and Strulovici, Bruno. Capital mobility and asset pricing. *Econometrica*, 80(6): 2469–2509, November 2012.

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

Duffie, Darrell, Gârleanu, Nicolae, and Pedersen, Lasse Heje. Valuation in over-the-counter markets. *Review of Financial Studies*, 20(6):1865–1900, 2007.

Frazzini, Andrea and Pedersen, Lasse Heje. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.

Greenwood, Robin, Hanson, Samuel G., and Stein, Jeremy C. A comparative-advantage approach to government debt maturity. *Journal of Finance*, 70(4):1683–1722, 2015.

Gromb, Denis and Vayanos, Dimitri. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics*, 66:361–407, 2002.

Haubrich, Joseph G. A brief history of bank capital requirements in the United States. *Economic Commentary*, 05:1–6, 2020.

He, Zhiguo and Krishnamurthy, Arvind. Intermediary asset pricing. *American Economic Review*, 103(2):732–770, 2013.

He, Zhiguo and Xiong, Wei. Rollover risk and credit risk. *Journal of Finance*, 67(2):391–430, 2012.

Holmstrom, Bengt and Tirole, Jean. Private and public supply of liquidity. *Journal of Political Economy*, 106(1):1–40, 1998.

Huang, Ji. Banking and shadow banking. *Journal of Economic Theory*, 178(C):124–152, 2018.

Huber Wang, Amy. Market power in wholesale funding: A structural perspective from the triparty repo market. Technical report, 2023.

Infante, Sebastian. Private money creation with safe assets and term premia. *Journal of Financial Economics*, 136(3):828–856, 2020.

Krishnamurthy, Arvind. The bond/old-bond spread. *Journal of Financial Economics*, 66(2): 463–506, 2002.

Krishnamurthy, Arvind and Li, Wenhao. The demand for money, near-money, and treasury bonds. *Review of Financial Studies*, forthcoming.

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

Lenel, Moritz. Safe assets, collateralized lending and monetary policy. Technical report, 2020.

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

Li, Ye. Public liquidity and intermediated liquidity. Technical report, 2022.

Malherbe, Frederic. Self-fulfilling liquidity dry-ups. *Journal of Finance*, 69(2):947–970, 2014.

Martin, Antoine, McAndrews, James, Palida, Ali, and Skeie, David. Federal reserve tools for managing rates and reserves. Staff Reports 642, Federal Reserve Bank of New York, 2019.

Merton, Robert C. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, 1969.

Myers, Stewart C. Determinants of corporate borrowing. *Journal of Financial Economics*, 5 (2):147–175, 1977.

Nagel, Stefan. The liquidity premium of near-money assets. *Quarterly Journal of Economics*, 131(4):1927–1971, November 2016.

Ordoñez, Guillermo. Sustainable shadow banking. *American Economic Journal: Macroeconomics*, 10(1):33–56, January 2018.

Patinkin, Don. *Money, Interest, and Prices: An Integration of Monetary and Value Theory.* Row Peterson, 1956.

Piazzesi, Monika and Schneider, Martin. Payments, credit and asset prices. Technical report, 2021.

Plantin, Guillaume. Shadow banking and bank capital regulation. *Review of Financial Studies*, 28(1):146–175, 2015.

Poole, William. Commercial bank reserve management in a stochastic model: Implications for monetary policy. *Journal of Finance*, 23(5):769–791, December 1968.

Pozsar, Zoltan and Sweeney, James. Global money notes #1: The money market under government control. Technical report, Credit Suisse, 2015.

Rotemberg, Julio J. Sticky prices in the United States. *Journal of Political Economy*, 90(6): 1187–1211, 1982.

Siriwardane, Emil, Sunderam, Adi, and Wallen, Jonathan L. Segmented arbitrage. Working Paper 30561, National Bureau of Economic Research, October 2022.

Stein, Jeremy C. Monetary policy as financial stability regulation. *Quarterly Journal of Economics*, 127(1):57–95, February 2012.

Stock, James and Yogo, Motohiro. Asymptotic distributions of instrumental variables statistics with many instruments. In *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, volume 6, pages 109–120. Cambridge University Press Cambridge, 2005.

Sunderam, Adi. Money creation and the shadow banking system. *Review of Financial Studies*, 28(4):939–977, 2015.

Tobin, James. Commercial banks as creators of "money". Technical report, Yale University, 1963.

Venkateswaran, Venky and Wright, Randall. Pledgability and liquidity: A new monetarist model of financial and macroeconomic activity. *NBER Macroeconomics Annual*, 28:227–270, 2013.

Wang, Jiang. A model of intertemporal asset prices under asymmetric information. *Review of Economic Studies*, 60(2):249–282, 1993.

Woodford, M. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, 2003. ISBN 9780691010496.

Xiao, Kairong. Monetary transmission through shadow banks. *Review of Financial Studies*, 33 (6):2379–2420, 2020.

Appendices

A Data Sources

We report data sources and variables used throughout this article. Yields are computed by taking the average of the daily observations within the month or week. Since money market instruments are quoted using different conventions, we convert all raw daily money market rates to effective annual yields using the 365.25 days-a-year convention. We use the CRSP US Treasury database and Daily Time Series file, and for each day select the Treasury Bill security with maturity closest to 91 days. We then calculate the yield from the midpoint of the bid and ask quotes reported in the CRSP.

- 1-month Treasury Bill Rate: We use the 4-Week Treasury Bill Secondary Market Rate (DTB4WK) available from FRED.
- Overnight Tri-Party Repo Rate: We collect the series for the Tri-Party General Collateral Rate (TGCR) from the Federal Reserve Bank of New York's website, which is available starting in August 2014. Before then, we retrieve from Bloomberg the USD Overnight GC Govt Repo series (USRG1T CMPN Curncy).
- IOR: We retrieve from FRED the Interest Rate on Excess Reserves series (IOER). Starting July 2021, the series was replaced with the Interest Rate on Reserve Balances (IORB).
- VIX: We use the monthly average VIX, which is available from FRED starting in January 1990.
- Weekly Treasury Bills Outstanding: We derive the amount of T-Bills outstanding from Treasury auctions. Precisely, we collect detailed data about the Auction Date, Offering Amount and Maturity Date from the published auction results available from the Treasury Direct website.
- T-Bills SOMA Holdings: T-bills SOMA Holdings are available at weekly frequency from the New York Fed's website starting in July 2003. We aggregate the amounts of regular Bills and Cash Management Bills and use the end-of-month values to align with total monthly T-bills outstanding. For the sample before 2003, we download yearly SOMA holdings data from Bukhari, Cambron, Fleming, McCarthy, and Remache (2013), who provide year-end historical SOMA holdings of Treasury bills over 1929-2012. We use linear interpolation to derive the monthly observations over the earlier sample before July 2003.
- Reverse Repo Volumes: We collect data about the Fed's RRP facility volumes from the New York Fed's website.
- Money Market Fund Holdings: We gather data from iMoneyNet's Money Fund Analyzer.

B Relegated Derivations

Guess and Verify We guess and verify that the value function of each agent has the following additive form:

$$V(\xi_t, n_t) = \xi_t + \frac{\log(n_t)}{\rho}, \qquad \overline{V}(\overline{\xi}_t, \overline{n}_t) = \overline{\xi}_t + \frac{\log(\overline{n}_t)}{\rho}, \qquad V^h\left(\xi_t^h, n_t^h\right) = \xi_t^h + \frac{\log(n_t^h)}{\rho},$$

for some stochastic processes $\{\xi_t, \overline{\xi}_t, \xi_t^h\}$ that capture time variations in the set of investment opportunities for a given type of agent. A unit of net worth has a higher value for a traditional bank, a shadow bank, or a household in states in which ξ_t, $\overline{\xi}_t$ or ξ_t^h are respectively high. We postulate that these wealth multipliers follow geometric Brownian motions.

Hamilton-Jacobi-Bellman Equation Using the guess and substituting for the balance sheet identity, the HJB equation for traditional banks can be written as

$$\rho\xi_t + \log(n_t) = \max_{\{w_t^b \geq 0, w_t^p, w_t^m \geq 0, w_t^d \geq 0, c_t \geq 0\}} \left\{ \log(c_t n_t) + \frac{\mu_t^n}{\rho} - \frac{1}{2}\frac{\psi_t^2}{\rho} + \mu_t^\xi \xi_t \right\}, \tag{34}$$

where

$$\begin{aligned}
\mu_t^n =& w_t^b(r_t^b - r_t^i) + w_t^m(r_t^m - r_t^i) + w_t^p(r_t^p - r_t^i) - w_t^d(r_t^d - r_t^i) - c_t - \mu_t^\tau \\
&- \chi\left([w_t^i]^+ + w_t^m + w_t^b + [w_t^p]^+\right), \\
\psi_t =& \lambda[\sigma^d w_t^d - \theta^m w_t^m - \theta^b w_t^b - \theta^p w_t^p]^+.
\end{aligned}$$

Shadow banks' and households' problems are nested by the one of traditional banks such that their respective HJB equations can be inferred from equation (34).

We derive the first-order conditions for the empirically relevant case: $w_t^p \leq 0$ and $w_t^i \geq 0$. Thus, $\chi([w_t^i]^+ + w_t^b + w_t^m + [w_t^p]^+)n_t = \chi(1 + w_t^d + \tau_t - w_t^p)n_t$. The first-order conditions for w_t^b, w_t^m, w_t^p, w_t^d and c_t are then given by equations (15) to (19).

C Proofs

For ease of notation, in this appendix, we denote Markov functions $g_t = g(x_t)$ as g whenever it is not confusing.

Proposition 1

If $\chi = 0$, then the first-order condition for repo of traditional banks in equation (18) becomes $r^i - r^p = \lambda\theta^p\psi$. Together with the first-order condition for repo of shadow banks in equation (22), we get $\psi = \overline{\psi}$.

Corollary 1

By definition, the return on a unit of illiquid risk-free capital is equal to $r^i = \rho$, given that the productivity of capital is equal to ρ and the price of capital is constant and equal to 1 (see footnote 6).

Thus, the first order conditions of traditional and shadow banks pin down all the other rates and only depend on individual controls through the couple $(\psi, \overline{\psi})$. Given Proposition 1, we have $\psi = \overline{\psi}$. Substitute in all market clearing conditions to find

$$\psi = \lambda \left[\gamma \sigma^d \frac{1 - n - \overline{n} + \tau^h}{n} - \theta^m \frac{m}{n} - \theta^b \frac{b - m - \overline{w}^b \overline{n}}{n} + \theta^p \frac{\overline{w}^p \overline{n}}{n} \right]^+, \tag{35}$$

$$\overline{\psi} = \lambda \left[(1-\gamma) \sigma^d \frac{1 - n - \overline{n} + \tau^h}{\overline{n}} - \theta^b \overline{w}^b - \theta^p \overline{w}^p \right]^+. \tag{36}$$

Thus, we have a linear system with 1 equation $\psi = \overline{\psi}$ and two variables $\overline{w}^b$ and $\overline{w}^p$, which is underdetermined.

Proposition 2

From Proposition 1, we have that $\psi(x) = \overline{\psi}(x)$. From the definition of $\psi(x)$ and the market-clearing condition for reserves, it directly follows that $\overline{\psi}(x; m^\star) = \psi(x; m^\star) > \psi(x; m^{\star\star}) = \overline{\psi}(x; m^{\star\star})$. Since $\psi(x; m^\star) > \psi(x; m^{\star\star})$ and $m^\star > 0$, using the first-order condition (17) yields that $r^i(x; m^{\star\star}) - r^m(x; m^{\star\star}) < r^i(x; m^\star) - r^m(x; m^\star)$. Since $b > 0$ and both $\psi(x; m^\star) > \psi(x; m^{\star\star})$ and $\overline{\psi}(x; m^\star) > \overline{\psi}(x; m^{\star\star})$, using the first-order conditions (16) and (21) yield that $r^i(x; m^\star) - r^b(x; m^\star) > r^i(x; m^{\star\star}) - r^b(x; m^{\star\star})$. Similarly, $r^i(x; m^\star) - r^p(x; m^\star) > r^i(x; m^{\star\star}) - r^p(x; m^{\star\star})$.

Lemma 1

We prove the lemma by contradiction and assume that $w^p < 0$. First notice that if $\mathcal{M}(x) \in \mathcal{I}$, then, by definition, $w^b > 0$ and $r^i - r^b = \lambda \theta^b \psi$. Thus, $\psi \geq \overline{\psi}$ since $r^i - r^b = \lambda \theta^b \psi \geq \lambda \theta^b \overline{\psi}$. Finally, $r^b - r^p = \lambda(\theta^p - \theta^b)\psi + \chi$ and $r^b - r^p \leq \lambda(\theta^p - \theta^b)\overline{\psi}$, which is a contradiction since $\psi \geq \overline{\psi}$ and $\chi > 0$.

Proposition 3

If $\mathcal{M}(x) \in \mathcal{S}$, then, by definition, $r^i - r^b > \lambda \theta^b \psi$. In that case, following the complementary slackness condition, $w^b = 0$ and since $b > 0$, we need that $\overline{w}^b(x) > 0$ to satisfy the market-clearing condition of T-bills. Thus, $r^i(x) - r^b(x) = \lambda \theta^b \overline{\psi}$ and $\psi < \overline{\psi}$. Moreover $\mathcal{M}(x) \in \mathcal{S}$ also implies by definition that $r^i - r^p < \lambda \theta^p \psi(x) + \chi$ and thus $w^p = \overline{w}^p = 0$. Shadow banks' first-order condition for repo then implies $r^i - r^p = \lambda \theta^p \overline{\psi} < \lambda \theta^p \psi + \chi$. Dividing by $\lambda \theta^p$ then yields the second inequality: $\overline{\psi} < \psi + \chi / (\lambda \theta^p)$.

Proposition 4

From Proposition 3, since $\mathcal{M}(x) \in \mathcal{S}$, we have that $\psi(x) < \overline{\psi}(x)$. Since $r^i - r^b > \lambda\theta^b\psi$ and $r^i - r^p < \lambda\theta^p\psi + \chi$, $w^b = w^p = \overline{w}^p = 0$. Since $\mathcal{M}(x) \notin \mathcal{E}$, $\psi(x) > 0$. From the definition of ψ and the market-clearing condition for reserves, it directly follows that $\psi(x; m^\star) > \psi(x; m^{\star\star})$. Using the market-clearing condition for T-bills and shadow bank deposits and the budget constraint of the central bank, we get $\overline{\psi}(x; m) = \lambda\left((1-\gamma)\sigma^d(1 - n - \overline{n} + \tau^h)/\overline{n} - \theta^b(b-m)/\overline{n}\right)$.

Thus, $\overline{\psi}(x; m^\star) < \overline{\psi}(x; m^{\star\star})$.

Since $\psi(x; m^\star) > \psi(x; m^{\star\star})$ and $m^\star > 0$, using the first-order condition (17) yields that $r^i(x; m^\star) - r^m(x; m^\star) > r^i(x; m^{\star\star}) - r^m(x; m^{\star\star})$.

Since $\overline{\psi}(x; m^\star) < \overline{\psi}(x; m^{\star\star})$ and $b > 0$, using the first-order condition (21) yields that $r^i(x; m^\star) - r^b(x; m^\star) < r^i(x; m^{\star\star}) - r^b(x; m^{\star\star})$.

Finally, since $\overline{\psi}(x; m^\star) < \overline{\psi}(x; m^{\star\star})$, using the first-order condition (22) yields that $r^i(x; m^\star) - r^p(x; m^\star) < r^i(x; m^{\star\star}) - r^p(x; m^{\star\star})$.

Proposition 5

If $m < m^{\mathcal{S}}$, then $\mathcal{M}(x) \notin \mathcal{S}$ and $\psi(x; m) = \overline{\psi}(x; m)$ from Proposition 1 so that Lemma 1 implies $w^p = \overline{w}^p = 0$. Thus, we can solve for $\overline{w}^b(x)$ in equations (35) and (36) and get

$$\psi(x; m) = \overline{\psi}(x; m) = \sigma^d \frac{1 - n - \overline{n} + \tau^h}{n + \overline{n}} - \theta^b \frac{b - m}{n + \overline{n}} - \theta^m \frac{m}{n + \overline{n}}.$$

If $m^{\mathcal{S}} < m < m^{\mathcal{E}}$, then $\mathcal{M}(x) \in \mathcal{S} \cap \mathcal{E}^c$. If $\mathcal{M}(x) \in \mathcal{S}$, $r^i(x) - r^b(x) > \lambda\theta^b\psi(x)$ and $r^i - r^p < \lambda\theta^p\psi(x) + \chi$. In that case, following the complementary slackness condition in equation (16) and (18), $w^b(x) = 0$ and $w^p(x) = 0$. Since $\mathcal{M}(x) \notin \mathcal{E}$, $\psi(x) > 0$. Thus, using the market-clearing condition for T-bills and reserves, we get

$$\begin{aligned}
\psi(x; m) &= \lambda\left(\sigma^d \frac{\gamma}{n}(1 - n - \overline{n} + \tau^h) - \theta^m \frac{m}{n}\right), \\
\overline{\psi}(x; m) &= \lambda\left(\sigma^d \frac{1-\gamma}{\overline{n}}(1 - n - \overline{n} + \tau^h) - \theta^b \frac{b - m}{\overline{n}}\right).
\end{aligned}$$

If $m > m^{\mathcal{E}}$, then $\mathcal{M}(x) \in \mathcal{S} \cap \mathcal{E}$. If $\mathcal{M}(x) \in \mathcal{E}$, then by definition $\psi(x) = 0$.

If $m > m^{\mathcal{A}}$, then $m > m^{\mathcal{E}}$ so that $\psi(x) = 0$. By definition $\mathcal{M}(x) \in \mathcal{A}$ implies $r^i(x) - r^b(x) > \lambda\theta^b\psi(x)$ and $w^p(x) < 0$. In that case, following the complementary slackness condition in equation, $w^b(x) = 0$. Using the first-order conditions (18) and (22), the following equality must hold: $r^i - r^p = \lambda\theta^p\overline{\psi}(x) = \lambda\theta^p\psi(x) + \chi$. Given $\psi(x) = 0$, we have $\overline{\psi}(x) = \chi/(\lambda\theta^p)$.

Lemma 2

If $\{m(x), i^m(x)\}$ is able to implement a given inflation target $\pi^\star$, that means that

$$i^m(x) = r^m(x;m) + \pi^\star.$$

For a combination of $\{m^\star(x), i^{m,\star}(x)\} \neq \{m(x), i^m(x)\}$ to be able to also implement $\pi^\star$, it is sufficient to show that there exists $m^\star \neq m$ such that

$$\left.\frac{\partial r^m(x;m)}{\partial m}\right|_{m=m^\star} \neq 0.$$

Indeed, if that partial derivative is everywhere equal to 0, then there is only a unique $i^m(x)$ that can implement $\pi^\star$. If that derivative is not everywhere equal to 0, because $i^m(x)$ is not restricted, we can find another $m(x)$ such that $r^m(x;m)$ and $i^m(x)$ are different and satisfies $i^m(x) = r^m(x;m) + \pi^\star$.

Given condition C, it is always possible to pick a $m^\star$ sufficiently small such that $\mathcal{M}(x) \in \mathcal{I}$; that is, $m^\star < m^{\mathcal{S}}$. Solving for the equilibrium prices, we find that

$$r^m = \rho - \lambda\theta^m \left(\sigma^d \frac{1-n-\overline{n}+\tau^h}{n+\overline{n}} - \theta^b \frac{b-m}{n+\overline{n}} - \theta^m \frac{m}{n+\overline{n}}\right),$$

and

$$\frac{\partial r^m(x;m)}{\partial m} = \lambda\theta^m \frac{\theta^m - \theta^b}{n+\overline{n}}.$$

Since $\frac{\partial r^m(x;m)}{\partial m}$ is constant, there exists a linear combination of $\{m^\star(x), i^{m,\star}(x)\} \neq \{m(x), i^m(x)\}$ that also implements $\pi^\star$.

Proposition 6

Since $\mathcal{M} \in \mathcal{I} \cap \mathcal{E}^c$, Proposition 5 implies that

$$\psi(x;m) = \overline{\psi}(x;m) = \lambda \left(\sigma^d \frac{1-n-\overline{n}+\tau^h}{n+\overline{n}} - \theta^m \frac{m}{n+\overline{n}} - \theta^b \frac{b-m}{n+\overline{n}}\right).$$

and

$$r^m(x;m) = \rho - \lambda\theta^m \psi(x;m),$$

The Fisher equation is given by

$$i^m = r^m(x;m) + \pi^*. \tag{37}$$

Given that i^m, π^*, and $\frac{\partial r^m(x;m)}{\partial m}$ are constant, there is only one m such that (37) is satisfied and $\psi(x;m)$ is constant as well. Thus equilibrium prices and allocations are constant when b changes.

Proposition 7

Since $\mathcal{M}(x) \in \mathcal{I} \cap \mathcal{E}$, we have that $\psi(x) = \overline{\psi}(x) = 0$. Thus, for any $b^\star$ and $b^{\star\star}$, $\psi(x; m, b^\star) = \psi(x; m, b^{\star\star}) = \overline{\psi}(x; m, b^\star) = \overline{\psi}(x; m, b^{\star\star})$. Hence, equilibrium allocations and prices are not affected.

Proposition 8

(i) When money markets are segmented and traditional banks are not liquidity satiated, $\psi(x; m) = \lambda\left(\sigma^d \gamma(1 - n - \overline{n} + \tau^h)/n - \theta^m m/n\right)$ and is not a function of b. When traditional banks are liquidity satiated, $\psi(x; m) = 0$. (ii) Hence, from $r^i(x; m) = \rho - \lambda\theta^m \psi(x; m)$ and equation (26), if the central bank does not change i^m or m, π does not change either. Consequently, given condition C, there exist some two (small) b^* and $b^{**}(> b^*)$ such that $\overline{\psi}(n, \overline{n}, b^*) > 0$ and $\overline{\psi}(n, \overline{n}, b^{**}) > 0$. Then from Proposition 4, $\overline{\psi}(n, \overline{n}, b) = \lambda\left(\sigma^d \frac{1-\gamma}{\overline{n}}(1 - n - \overline{n} + \tau^h) - \theta^b \frac{b-m}{\overline{n}}\right)$ so that $\overline{\psi}(n, \overline{n}, b^*) > \overline{\psi}(n, \overline{n}, b^{**})$.

D New-Keynesian Extension

In this appendix, we solve an extension of the model with a continuum of firms in monopolistic competition using labor to produce intermediate goods and facing price adjustment costs as in Rotemberg (1982). The goal is to derive the New Keynesian investment-saving and Philips curves. To simplify the exposition, we abstract from aggregate risk.

Below, we solve the extended problem of an agent with utility for leisure. The solution can be immediately applied to households, traditional banks, and shadow banks by adding the portfolio problem specific to each agent described in the main text.

$$\max_{\{c_u, \ell_u\}_{u=t}^{\infty}} E_t \left[\int_t^\infty e^{-\rho(u-t)} \left(\log(c_u) + \varphi \log(1 - \ell_u)\right) du \right],$$

subject to the law of motion of wealth:

$$dn_t = \left(r_t^i n_t + \ell_t w_t - c_t + \mu_t^\tau n_t\right) dt.$$

Deriving the HJB with the value function given by $V(\xi_t, n_t) = \xi_t + \frac{\log(n_t)}{\rho}$ and taking the first-order conditions for consumption c_t and labor ℓ_t yields

$$c_t = \rho n_t,$$
$$\frac{\varphi}{1 - \ell_t} = \frac{w_t}{\rho n_t}.$$

A competitive final goods producer aggregates a continuum of intermediate goods with nominal

price p_t^j:

$$c_t = \left(\int_0^1 (c_t^j)^{\frac{\zeta-1}{\zeta}} dj \right)^{\frac{\zeta}{\zeta-1}}.$$

Minimization of costs yields the following demand for intermediate good j:

$$c_t^j = \left(\frac{p_t^j}{p_t} \right)^{-\zeta} c_t$$

where

$$p_t = \left(\int_0^1 (p_t^j)^{1-\zeta} dj \right)^{\frac{1}{1-\zeta}}$$

Ownership of capital gives right to the dividends of the continuum of monopolistically competitive intermediate goods producers. Production uses labor only according to:

$$y_t^j = a\ell_t^j.$$

The firms aim to set prices to maximize the value of future profits but have to pay quadratic price adjustment cost:

$$\frac{\theta}{2} \left(\frac{\dot{p}_t^j}{p_t^j} \right)^2,$$

where we use the over-dot to represent the time derivative: $\dot{x} = dx_t/dt$. The optimal control problem is given by:

$$\Pi(p_0^j; \mathbf{x}_0) = \max_{p_t^j} \int_0^\infty e^{-\int_0^t i_s^i ds} \left(p_t^j y_t^j - p_t w_t \ell_t^j - \frac{\theta}{2} \left(\frac{\dot{p}_t^j}{p_t^j} \right)^2 p_t y_t \right) dt$$

where $p_t w_t$ is the nominal wage and $i_t^i = r_t^i + \pi_t$ is the nominal illiquid interest rate and inflation is given by $\pi_t \equiv \frac{\dot{p}_t}{p_t}$. The HJB can be written as

$$i_t^i \Pi(p_t^j, x) = \max_{\pi_t^j} \left\{ p_t^j y_t^j - p_t w_t \ell_t^j - \frac{\theta}{2} (\pi_t^j)^2 p_t y_t + \pi_t^j p_t^j \frac{\partial \Pi_t}{\partial p_t^j} + \boldsymbol{\mu}_t^x x \frac{\partial \Pi_t}{\partial x} \right\},$$

where x_t represents the state of the economy. Note that if $\theta = 0$, the firm can directly maximize on p_t^j and we get $p_t^j = \frac{\zeta}{\zeta-1} \frac{p_t w_t}{a}$. Using the envelope theorem to solve for the partial derivatives of Π_t and assuming a symmetric equilibrium such that $p_t^j = p_t$, we obtain a law of motion for inflation:

$$\pi_t \left(i_t^i - \pi_t - \frac{\dot{y}_t}{y_t} \right) = \frac{\zeta - 1}{\theta} \left(\frac{\zeta}{\zeta - 1} \frac{w_t}{a} - 1 \right) + \dot{\pi}_t.$$

Without risk, the nominal total return (dividends plus capital gains) on the nominal net worth $p_t n_t$ of the economy needs to be equal to the illiquid nominal rate. Thus,

$$i_t^i = \frac{a\ell_t}{n_t} + \frac{\dot{n}_t}{n_t} + \pi_t = \rho + \frac{\dot{y}_t}{y_t} + \pi_t,$$

where the last equality comes from the market clearing condition for the consumption good: $p_t c_t = \rho n_t = p_t y_t = p_t a \ell_t$. Thus, we get

$$\frac{\dot{y}_t}{y_t} = i_t^i - \pi_t - \rho,$$

which is the New Keynesian investment-saving curve. Using the first-order condition for labor to substitute for w_t/a,

$$\rho \pi_t = \frac{\zeta - 1}{\theta} \left(\frac{\zeta}{\zeta - 1} \frac{\varphi y_t}{a - y_t} - 1 \right) + \dot{\pi}_t,$$

which is the New Keynesian Phillips curve. Note that in this environment, the Fisher equation is unchanged and still given by

$$i_t^m = r_t^m + \pi_t$$

and the relationship between the nominal interest rate on reserves and the nominal interest rate on the illiquid capital is still given by

$$i_t^i = i_t^m + \lambda \theta^m \psi_t.$$

E Econometric Model

We assume that we are in a regime of both segmentation and bank liquidity satiation, $\mathcal{M}(x) \in \mathcal{S} \cap \mathcal{E}$, and that the balance sheet cost χ is higher than the reverse repo facility rate spread such that traditional banks never have incentives to intermediate liquidity to shadow banks. As mentioned in the main text, we treat the wealth shares of the traditional and shadow banking sectors as constants. The model yields three key equations. First, if $f_t \leq \underline{f}$, we have that

$$r_t^m - r_t^b = \lambda^2 \theta^b \left(\sigma_t^d (1 - \gamma) \frac{1 - n - \overline{n} + \tau^h}{\overline{n}} - \theta^b \frac{b_t^\star}{\overline{n}} \right), \tag{38}$$

where $b_t^\star$ = T-bills Outstanding$_t$ - $\beta\times$ Govt. MMF$_t$ and $\sigma_t^d = \sigma^d + \sigma_c \times \mathbb{1}\{t \text{ after January 1st, 2021}\}$. This equation can be re-written as

$$r_t^m - r_t^b = \underbrace{\tilde{\theta}^b \left(\tilde{\sigma}_t^d - \tilde{\theta}^b b_t^\star\right)}_{x_t^b}, \tag{39}$$

where $\tilde{\theta}^b \equiv \lambda\theta^b/\sqrt{\overline{n}}$, $\tilde{\sigma}^d \equiv \lambda\sigma^d(1-\gamma)(1-n-\overline{n}+\tau^h)/\sqrt{\overline{n}}$, $\tilde{\sigma}^c \equiv \lambda\sigma^c(1-\gamma)(1-n-\overline{n}+\tau^h)/\sqrt{\overline{n}}$, $\tilde{\sigma}_t^d = \tilde{\sigma}^d + \tilde{\sigma}_c \times \mathbb{1}\{t \text{ after January 1st, 2021}\}$. When repo rates are below that of the repo facility, the Fed absorbs all adjustments in liquidity risk such that $\overline{\psi}_t \geq \overline{\psi}_t^\star$, where $\overline{\psi}_t^\star$ is defined as the quantity of shadow bank liquidity risk when repo rates are exactly equal to the repo facility rate $r_t^p = r_t^f$. That is, $\overline{\psi}_t^\star \equiv (r_t^m - r_t^f)/(\lambda\theta^p)$. Thus, if $f_t > \underline{f}$, we get that

$$r_t^m - r_t^b = r_t^m - r_t^f + \lambda(\tilde{\theta}^b - \tilde{\theta}^p)\overline{\psi}_t^\star. \tag{40}$$

where $\tilde{\theta}^p \equiv \lambda\theta^p/\sqrt{\overline{n}}$. Similarly, if $f_t \leq \underline{f}$, the second equation implied by the model is given by

$$r_t^m - r_t^p = \underbrace{\tilde{\theta}^p(\tilde{\sigma}_t^d - \tilde{\theta}^b b_t^\star)}_{x_t^p}, \tag{41}$$

and, if $f_t > \underline{f}$,

$$r_t^m - r_t^p = r_t^m - r_t^f. \tag{42}$$

Finally, the RRP facility volumes necessary to keep the liquidity of the shadow banking sector bounded at $\overline{\psi}_t^\star$ is given by

$$f_t = \underbrace{\frac{\tilde{\sigma}_t^d - \tilde{\theta}^b b_t^\star}{\tilde{\theta}^p} - \frac{r_t^m - r_t^f}{(\tilde{\theta}^p)^2}}_{x_t^f}. \tag{43}$$

Maximum Likelihood Estimation We assume that the error term between the prediction of the model and the observed spreads and volumes at the RRP facility are identically and independently distributed according to a Normal distributions with mean 0. That is,

$$r_t^m - r_t^b = x_t^b + \varepsilon_t^b \qquad \varepsilon_t^b \sim \mathcal{N}(0, (\sigma^b)^2), \tag{44}$$

$$r_t^m - r_t^p = x_t^p + \varepsilon_t^p \qquad \varepsilon_t^p \sim \mathcal{N}(0, (\sigma^p)^2), \tag{45}$$

$$f_t = x_t^f + \varepsilon_t^f \qquad \varepsilon_t^f \sim \mathcal{N}(0, (\sigma^f)^2). \tag{46}$$

Furthermore, we censor the spreads $r_t^m - r_t^b$ and $r_t^m - r_t^p$ above 0. Thus, given $\mathbf{y}_t = \{r_t^m, r_t^b, r_t^p, f_t\}$ and the set of parameters $\boldsymbol{\theta} = \{\tilde{\sigma}^d, \tilde{\sigma}^c, \tilde{\theta}^b, \tilde{\theta}^p, \beta, \underline{f}, \sigma^b, \sigma^p, \sigma^f\}$, the log-likelihood function is given

by

$$\begin{aligned}
\ln \mathcal{L}(\mathbf{y}_t;\boldsymbol{\theta}) =& I_t^p(1-I_t^f)\log\phi\left(\varepsilon_t^p,0,\sigma^p\right)+(1-I_t^p)(1-I_t^f)\log\Phi\left(0,x_t^p,\sigma^p\right) && (47)\\
&+I_t^f\log\left(1-\Phi\left(r_t^m-r_t^f,x_t^p,\sigma^p\right)\right) && (48)\\
&+I_t^b(1-I_t^f)\log\phi(\varepsilon_t^b,0,\sigma^b)+(1-I_t^b)(1-I_t^f)\log\Phi\left(0,x_t^b,\sigma^b\right) && (49)\\
&+I_t^f\log\left(1-\Phi\left(r_t^m-r_t^f+(\tilde{\theta}^b-\tilde{\theta}^p)(r_t^m-r_t^f)/\tilde{\theta}^p,x_t^b,\sigma^b\right)\right) && (50)\\
&+I_t^f\log\phi(\varepsilon_t^f,0,\sigma^f)+(1-I_t^f)\log\Phi\left(\underline{f},x_t^f,\sigma^f\right), && (51)
\end{aligned}$$

where $I_t^p=1$ if $r_t^m-r_t^p>0$ and 0 otherwise, $I_t^b=1$ if $r_t^m-r_t^b>0$ and 0 otherwise, $I_t^f=1$ if $f_t>\underline{f}$ and 0 otherwise, and ϕ and Φ are the probability density function and cumulative density function of the standard normal distribution. We set $\underline{f}$ to \$15 billion. The rest of the parameters are estimated by maximizing the likelihood function.

In the unrestricted estimation, equations (44), (45), and (46) are estimated independently with their own set of parameters and log-likelihood functions. We set the censoring of the $r_t^m-r_t^b$ at $r_t^m-r_t^f+0.0003$ instead of $r_t^m-r_t^f+(\tilde{\theta}^b-\tilde{\theta}^p)(r_t^m-r_t^f)/\tilde{\theta}^p$ to keep the estimation of each equation independent.

To estimate the standard errors of the parameters, we resample with replacement the original set of 670 observations $\mathbf{y}_t$ to create 2,000 random samples of 670 observations and rerun the maximum likelihood estimation on each of these samples.

Counterfactuals We construct each countefactual time series as the model's prediction plus the error term estimated in the benchmark estimation. As an example, the counterfactual time series of RRP volumes $\hat{f}_t$ such that the repo spreads $r_t^m-r_t^p$ is equal to 0 is given by

$$\hat{f}_t=\frac{\tilde{\sigma}_t^d-\tilde{\theta}^b\tilde{b}_t}{\tilde{\theta}^p}+\varepsilon_t^f, \qquad \text{where} \quad \varepsilon_t^f=f_t-\frac{\tilde{\sigma}_t^d-\tilde{\theta}^b\tilde{b}_t}{\tilde{\theta}^p}+\frac{r_t^m-r_t^f}{(\tilde{\theta}^p)^2}. \tag{52}$$

F Deposit Channel Extension

In this section, we allow households to choose among three different liquid assets: cash, traditional bank deposits, and shadow bank deposits. The problem of the households becomes:

$$\max_{\{c_u\geq 0,i_u\geq 0,m_u\geq 0\}_{u=t}^{\infty}} E_t\left[\int_t^{\infty} e^{-\rho(u-t)}U_u^h du\right], \tag{53}$$

subject to the law of motion of wealth:

$$dn_t=\left(i_t^h r_t^i+d_t^h r_t^d+\overline{d}_t^h\overline{r}_t^d-m_t^h\pi_t-\mu_t^{h,\tau}-c_t^h\right)dt \tag{54}$$

and the balance sheet constraint:

$$i_t^h + d_t^h + \overline{d}_t^h + m_t^h = n_t^h + \tau_t^h. \tag{55}$$

The representative household maximizes utility over consumption c_t^h and liquidity services ℓ_t^h according to a log-CES aggregator:

$$U_t^h = \log\left(z_u^h\right) = \frac{\varrho}{\varrho - 1}\log\left(\varpi^{\frac{1}{\varrho}}(c_t^h)^{\frac{\varrho-1}{\varrho}} + (1-\varpi)^{\frac{1}{\varrho}}\left(\ell_t^h\right)^{\frac{\varrho-1}{\varrho}}\right). \tag{56}$$

Liquidity services ℓ_t^h are themselves derived from holding traditional bank deposits, d_t^h, and other liquid assets x_t^h, also according to a CES aggregator:

$$\ell_t^h = \left(\alpha^{\frac{1}{\varepsilon}}(d_t^h)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\alpha)^{\frac{1}{\varepsilon}}(x_t^h)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{57}$$

with x_t^h also a CES aggregator between cash and shadow bank deposits:

$$x_t^h = \left(\gamma^{\frac{1}{\zeta}}(m_t^h)^{\frac{\zeta-1}{\zeta}} + (1-\gamma)^{\frac{1}{\zeta}}(\overline{d}_t^h)^{\frac{\zeta-1}{\zeta}}\right)^{\frac{\zeta}{\zeta-1}}. \tag{58}$$

Taking the first-order conditions, we get the usual solution with CES aggregators:

$$m_t^h = \gamma x_t^h\left(\frac{\kappa_t^x}{\kappa_t^m}\right)^{\zeta}, \quad \overline{d}_t^h = (1-\gamma)x_t^h\left(\frac{\kappa_t^x}{\overline{\kappa}_t^d}\right)^{\zeta}, \quad d_t^h = \alpha\ell_t^h\left(\frac{\kappa_t^\ell}{\kappa_t^d}\right)^{\varepsilon}, \tag{59}$$

$$x_t^h = (1-\alpha)\ell_t^h\left(\frac{\kappa_t^\ell}{\kappa_t^x}\right)^{\varepsilon}, \quad \ell_t^h = (1-\varpi)z_t^h\left(\frac{\kappa_t^z}{\kappa_t^\ell}\right)^{\varrho}, \quad c_t^h = \varpi z_t^h(\kappa_t^z)^{\varrho}, \quad z_t^h = \frac{\rho}{\kappa_t^z}n_t^h, \tag{60}$$

where the κ's are defined as the spreads with the illiquid rate:

$$\kappa_t^m \equiv r_t^i + \pi_t, \quad \overline{\kappa}_t^d \equiv r_t^i - \overline{r}_t^d, \quad \kappa_t^d \equiv r_t^i - r_t^d, \quad \kappa_t^x \equiv \left(\gamma(\kappa_t^m)^{1-\zeta} + (1-\gamma)(\overline{\kappa}_t^d)^{1-\zeta}\right)^{\frac{1}{1-\zeta}}, \tag{61}$$

$$\kappa_t^\ell \equiv \left(\alpha(\kappa_t^d)^{1-\varepsilon} + (1-\alpha)(\kappa_t^x)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}, \quad \kappa_t^z \equiv \left(\varpi + (1-\varpi)(\kappa_t^\ell)^{1-\varrho}\right)^{\frac{1}{1-\varrho}}. \tag{62}$$

We can then compute the demand own rate elasticities as:

$$-\frac{\kappa_t^d}{d_t^h}\frac{\partial d_t^h}{\partial \kappa_t^d} = \varepsilon + \left(\varrho - \varepsilon + (1-\varrho)(1-\varpi)\left(\frac{\kappa_t^\ell}{\kappa_t^z}\right)^{1-\varrho}\right)\alpha\left(\frac{\kappa_t^d}{\kappa_t^\ell}\right)^{1-\varepsilon} \tag{63}$$

$$-\frac{\overline{\kappa}_t^d}{\overline{d}_t^d}\frac{\partial \overline{d}_t^d}{\partial \overline{\kappa}_t^d} = \zeta - (1-\gamma)\left(\frac{\kappa_t^x}{\overline{\kappa}_t^d}\right)^{\zeta-1} \tag{64}$$

$$\times\left(\zeta + \left((\varrho-1)(1-\varpi)\left(\frac{\kappa_t^z}{\kappa_t^\ell}\right)^{\varrho-1} - \varrho + \varepsilon\right)(1-\alpha)\left(\frac{\kappa_t^\ell}{\kappa_t^x}\right)^{\varepsilon-1} - \varepsilon\right). \tag{65}$$

Estimation Given $\{b_t^\star, r_t^m, f_t\}$, we solve for d_t^h, $\overline{d}_t^h$, r_t^d, and $\overline{r}_t^d$ that clear the markets for deposits in every period t using the above system of equations as well as the two first-order conditions introduced in equations (32) and (33):

$$-\frac{\kappa_t^d}{d_t^h}\frac{\partial d_t^h}{\partial \kappa_t^d} = \frac{\kappa_t^d}{\chi - \kappa_t^d}, \tag{66}$$

$$-\frac{\overline{\kappa}_t^d}{\overline{d}_t^h}\frac{\partial \overline{d}_t^h}{\partial \overline{\kappa}_t^d} = \frac{\overline{\kappa}_t^d}{\lambda\sigma^d\overline{\psi}_t + \overline{\chi} - \overline{\kappa}_t^d}. \tag{67}$$

We then add two equations to the econometric model

$$r_t^m - r_t^d = \kappa_t^d + \varepsilon_t^d \qquad \varepsilon_t^d \sim \mathcal{N}(0, (\sigma^{rd})^2), \tag{68}$$

$$r_t^m - \overline{r}_t^d = \overline{\kappa}_t^d + \overline{\varepsilon}_t^d \qquad \overline{\varepsilon}_t^d \sim \mathcal{N}(0, (\overline{\sigma}^{rd})^2) \tag{69}$$

and modify the log-likelihood function accordingly. Using a maximum likelihood estimation procedure with the constraint that $\chi \geq 25$ basis points, our estimates are given by $\boldsymbol{\theta} = \{\hat{\sigma}^d : 3.1321 \times 10^{-7}, \hat{\sigma}^c : 3.3239 \times 10^{-7}, \tilde{\theta}^b : 0.254, \tilde{\theta}^p : 0.211, \beta : -0.324, \varrho : 2.036, \varepsilon : 3.1745, \zeta : 1.6803, \alpha : 0.9495, \gamma : 0.9495, \varpi : 0.499, \chi : 0.0025021, \overline{\chi} : 0.00099931, \tilde{n}^h : 101.87 \times 10^6, \sigma^b = 0.0006, \sigma^p = 0.0005, \sigma^f = 179.8470, \sigma^{rd} = 0.0031, \overline{\sigma}^{rd} = 0.0014\}$, where

$$\hat{\sigma}^d \equiv \lambda\sigma^d/\sqrt{\overline{n}}, \tag{70}$$

$$\hat{\sigma}^c \equiv \lambda\sigma^c/\sqrt{\overline{n}}, \tag{71}$$

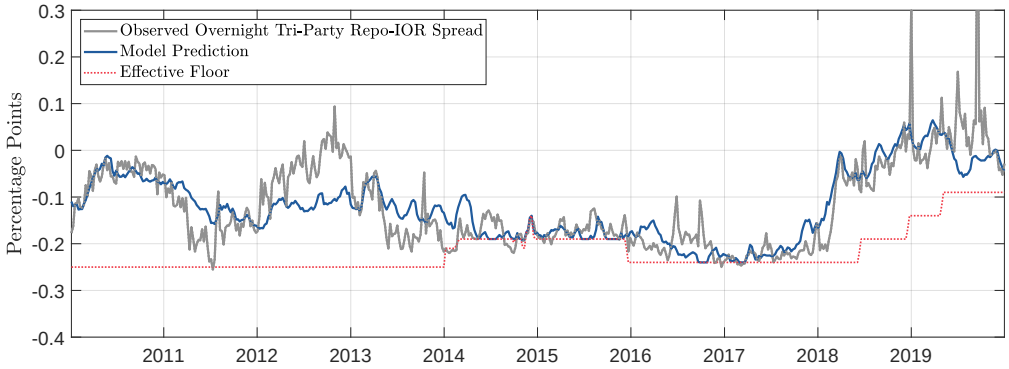
$$\tilde{n}^h \equiv (1-\varpi)\rho n^h. \tag{72}$$

Thus, the corresponding $\tilde{\sigma}^d$ and $\tilde{\sigma}^c$ of the benchmark model without the deposit channel are given by:

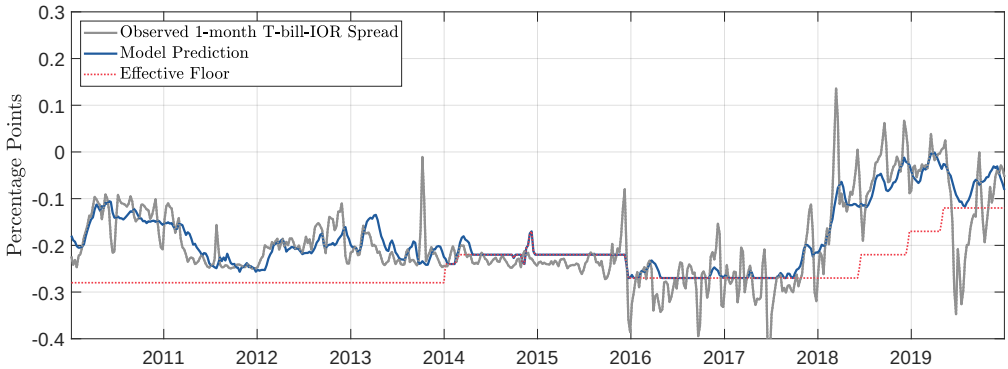
$$\tilde{\sigma}^d = \hat{\sigma}^d\mathbb{E}\left[\overline{d}_t^h\right] = 409.587, \tag{73}$$

$$\tilde{\sigma}^c = \hat{\sigma}^c\mathbb{E}\left[\overline{d}_t^h\right] = 434.671. \tag{74}$$

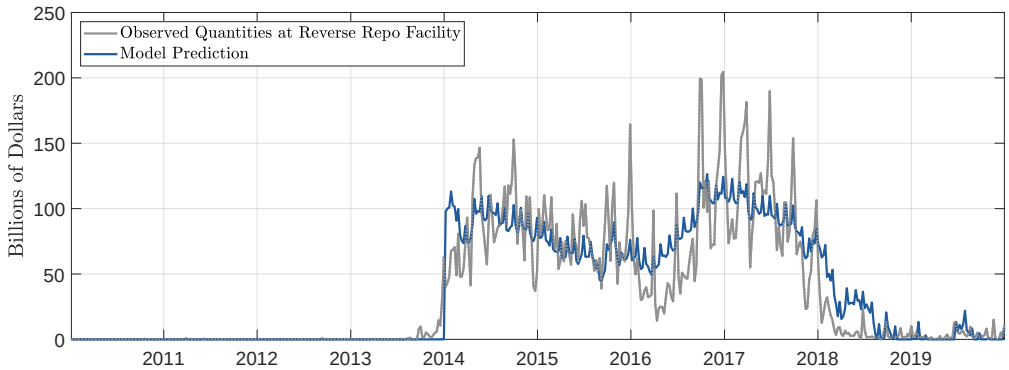
G Relegated Figures and Tables



(a) Model Predictions for the Spread between the Repo Rate and the IOR



(b) Model Predictions for the Spread between the T-bill yield and the IOR



(c) Model Predictions for Volumes at the Reverse Repo Facility

Figure 13: Predicted Values and Observations This figure shows the model predictions for the Tobit model from Table 1.

Table 3: Baseline Regressions with Reserves. For each triplet group, the table presents results from weekly regressions of 1-month T-bill (T-bill-IOR) and Overnight Tri-Party (TGCR-IOR) yield spreads in excess of IOR and of Quantities at the Reverse Repo Facility (RRP Vol.) on outstanding supply of Treasury bills (T-bill Supply), government-only money market funds assets under management (Govt MMFs), the interest on reserves (IOR), and Reserves. Columns (1)-(3) show results from a Tobit regression in a level specification. Control variables include: the VIX, a linear trend, and an end-of-month calendar fixed effect. Columns (4)-(6) report OLS estimates where all regression variables are expressed in a 4-week difference. Columns (7)-(9) present results from GMM regressions with all variables expressed in a 4-week difference. The endogenous variables are ΔT-bill Supply, ΔGovt MMFs, and Reserves and are instrumented using week-of-year polynomial dummies up to 18th degree, an instrumental variable based on the Prime-to-Government MMF Reform weekly TNA flow, and Treasury General Account balances. HAC adjusted standard errors, and GMM weighting matrix for the 4-week difference regressions are estimated using the Newey-West estimator with 12 lags, and Control variables include: the VIX and beginning- and end-of-month calendar fixed effects. The table row Sample denotes the regression estimation sample period split by the volumes at the RRP facility. The Cragg-Donald Wald F-statistic for the GMM procedure is 6.6.

	Level Tobit			4-Week Diff OLS			4-Week Diff GMM		
	(1) T-bill-IOR	(2) TGCR-IOR	(3) RRP Vol.	(4) ΔT-bill-IOR	(5) ΔTGCR-IOR	(6) ΔRRP Vol.	(7) ΔT-bill-IOR	(8) ΔTGCR-IOR	(9) ΔRRP Vol.
T-bill Supply	0.040***	0.020***	−0.143***	0.028**	0.024**	−0.291***	0.019*	0.024***	−0.220***
	(0.002)	(0.002)	(0.018)	(0.011)	(0.007)	(0.063)	(0.008)	(0.005)	(0.048)
Govt MMFs	−0.012***	−0.022***	0.120***	−0.035*	−0.026*	0.244***	−0.022	−0.030*	0.320***
	(0.001)	(0.001)	(0.008)	(0.016)	(0.012)	(0.066)	(0.013)	(0.013)	(0.065)
IOR	0.659	−0.160	15.23*	−28.72***	1.505	121.3***	−25.35***	0.383	163.0***
	(1.048)	(1.002)	(6.601)	(5.219)	(3.305)	(36.34)	(5.132)	(2.175)	(42.58)
Reserves	0.000	−0.015***	−0.066***	−0.007	−0.011**	−0.063*	0.015	−0.006	0.072
	(0.001)	(0.001)	(0.012)	(0.005)	(0.004)	(0.026)	(0.008)	(0.006)	(0.070)
Intercept	−83.68***	−78.69***	1377.***	0.558	0.249	−7.486	0.509	−0.038	−8.532*
	(5.881)	(5.386)	(134.6)	(0.418)	(0.292)	(3.906)	(0.392)	(0.241)	(3.406)
# Obs.	522	522	261	304	304	218	304	304	218
Period	2010-2019	2010-2019	2015-2019	2010-2019	2010-2019	2013-2019	2010-2019	2010-2019	2013-2019
Sample	Full	Full	Full	≤ $15bn.	≤ $15bn.	> $15bn.	≤ $15bn.	≤ $15bn.	> $15bn.
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

*p<0.05, ** p<0.01, *** p<0.001

Online Appendix

A Micro-Foundations for Liquidity Management

We describe the liquidity management problem of banks as a discrete-time problem with an interim period in which assets can only be traded at some cost. Then, we show that this problem converges to the continuous-time problems of traditional and shadow banks from Section 2. This micro-foundation draws inspiration from Bianchi and Bigio (2022) (adding fire sales and liquid money market asset holdings) and He and Xiong (2012) (adding reserves and liquid money market asset holdings). This micro-foundation is also similar to that of d'Avernas, Vandeweyer, and Darracq-Pariès (2019) when adding T-bills and repo transactions and not allowing for interbank trade during the illiquid stage.

Timing Time is discrete with an infinite horizon. Each period is divided into two stages: the liquid stage ℓ and the illiquid stage i. Both stages last a period of time Δt. In the liquid stage, there is no liquidity friction and portfolios can be adjusted at market prices without any cost. Then, the macroeconomic shock on risky securities realizes and interest rates are paid. At the beginning of the illiquid stage, deposits are randomly reshuffled from some banks—*the deficit banks*—to others—*the surplus banks*. Deficit banks cannot contract new loans and have to rely on disbursing existing assets in order to settle their debts with the surplus banks. There are two types of liquidity frictions in the illiquid stage. First, only a fraction of assets can be mobilized to settle debts. Second, it is costly to use assets during the illiquid stage for settlement purposes. This cost depends on the liquidity of the assets, with risky securities being the most illiquid. After the end of the illiquid stage, the economy enters a new liquid stage for the next period.

The Liquid Stage In the liquid stage, all banks can trade assets without friction. The law of motion for the wealth of banks in the liquid stage can then be written as

$$\Delta^{\ell} n_t = \Big(r_t^m m_t + r_t^i i_t + r_t^b b_t + r_t^p p_t - r_t^d d_t - c_t n_t + \mu_t^{\tau} n_t \Big) \Delta t. \tag{75}$$

Bankers face a portfolio choice problem with four different assets: securities portfolio s_t, treasury bills b_t, central bank reserves m_t, interbank lending i_t, and deposits d_t. In equation (75), r_t^i is the return on an illiquid asset, r_t^m the interest rate paid by the central bank on its reserves, r_t^b the interest rate paid by the government on T-bills, r_t^p the interest rate on repos, and r_t^d the interest rate on deposits. Banks also choose their consumption rate c_t as a fraction of their wealth and receive a flow of transfers per unit of wealth of μ_t^{τ}.

The Illiquid Stage Each individual bank is subject to an *idiosyncratic* deposit shock:

$$\Delta^i d_t = \sigma_t^d d_t \varepsilon_t^i \sqrt{\Delta t},$$

where ε_t^i is a binomial stochastic variable distributed with even probabilities:

$$\varepsilon_t^i = \begin{cases} +1 & \text{with } p = 1/2, \\ -1 & \text{with } p = 1/2. \end{cases}$$

In the illiquid period, interbank loans i_t cannot be contracted. The balance sheet constraint of the bank imposes that the flow of deposits is matched with an equivalent flow of securities. That is,

$$\Delta^i m_t + \Delta^i p_t + \Delta^i i_t + \Delta^i b_t = \Delta^i d_t.$$

The flows of assets $\Delta^i s_t$, $\Delta^i i_t$, $\Delta^i p_t$, $\Delta^i m_t$, and $\Delta^i b_t$ are chosen by deficit banks in order to minimize the net cost of transactions. To simplify the model, we assume that the costs of trading illiquid assets are fixed exogenously[22] and transferred from deficit to surplus banks. We capture these costs with parameters λ^s, λ^m, λ^p, λ^i, and λ^b. Surplus banks do not face liquidity constraints and take these opportunities to purchase these assets at a discounted price as given. Because the policy functions are linear in the agents' wealth, the distribution of these flows does not impact the recursive competitive equilibrium.

We can then write the net impact of the cost of the deposit shock on an individual bank's wealth as

$$\Delta^i n_t = \lambda^i \Delta^i i_t + \lambda^m \Delta^i m_t + \lambda^p \Delta^i p_t + \lambda^b \Delta^i b_t.$$

Substituting for the balance sheet constraint, we have

$$\Delta^i n_t = \lambda^m \Delta^i m_t + \lambda^p \Delta^i p_t + \lambda^b \Delta^i b_t + \lambda^i \left(\Delta^i d_t - \Delta^i m_t - \Delta^i p_t - \Delta^i b_t \right),$$

which can be rewritten as

$$\Delta^i n_t = \lambda^i \left(\Delta^i d_t - \frac{\lambda^i - \lambda^m}{\lambda^i} \Delta^i m_t - \frac{\lambda^i - \lambda^p}{\lambda^i} \Delta^i p_t - \frac{\lambda^i - \lambda^b}{\lambda^i} \Delta^i b_t \right). \tag{76}$$

Moreover, a second type of liquidity friction constrains the number of assets that can be sold by deficit banks during the time interval Δt. A deficit bank can only decrease its asset holdings and only up to a certain threshold. In order to converge to a Brownian motion in the continuous time approximation, this amount is proportional to $\sqrt{\Delta t}$. For example, a deficit bank cannot sell more than a fraction $\delta^s \sqrt{\Delta t}$ of its risky securities over the interval Δt. We write these

[22] We do not provide a micro-foundation for the cost of a fire sale, but we refer to the large literature in which it arises either as a consequence of the shift in bargaining power under strong selling pressure (see Brunnermeier and Pedersen, 2005; Duffie and Strulovici, 2012; Duffie, Gârleanu, and Pedersen, 2005, 2007) or asymmetry of information (see Malherbe, 2014; Wang, 1993). The intuition is that using reserves or other liquid money market assets has a negligible cost compared with having to sell risky securities. The intuition for including short-maturity loans as liquid assets is that if the illiquid stage lasts for a longer period than the maturity of the short-term loan, the bank will be able to use the funds lent at the due date, thereby creating a liquidity component of the term structure as modeled by Acharya and Skeie (2011) and documented empirically by Greenwood et al. (2015).

constraints as

$$0 \geq \Delta^i i_t \geq -\delta^i i_t \sqrt{\Delta t}, \tag{77}$$
$$0 \geq \Delta^i m_t \geq -\delta^m m_t \sqrt{\Delta t}, \tag{78}$$
$$0 \geq \Delta^i p_t \geq -\delta^p p_t \sqrt{\Delta t}, \tag{79}$$
$$0 \geq \Delta^i b_t \geq -\delta^b b_t \sqrt{\Delta t} \tag{80}$$

The optimization problem of deficit banks in the illiquid stage amounts to the static[23] minimization of their losses under the liquidity constraints

$$\min_{\Delta^i p_t, \Delta^i m_t, \Delta^i i_t, \Delta^i b_t} \Delta^i n_t$$

where $\Delta^i n_t$ is given by (76), $\Delta^i d_t = -\sigma_t^d \sqrt{\Delta t}$ and is subject to previously stated liquidity frictions.

We first consider the case in which liquid assets are not sufficient for a deficit bank to cover its funding needs; that is, $\sigma_t^d d_t > \delta^m m_t + \delta^p p_t + \delta^b b_t$. Since using illiquid assets i_t is the most costly asset, deficit banks always first use their liquid assets m_t, b_t, and p_t and only then resort to selling securities in order to settle remaining due debt positions. Hence, the optimal portfolio adjustments are given by

$$\begin{aligned}
\Delta^i i_t &= \Delta^i d_t + \Delta^i m_t + \Delta^i p_t + \Delta^i b_t, \\
\Delta^i m_t &= -\delta^m m_t \sqrt{\Delta t}, \\
\Delta^i p_t &= -\delta^i p_t \sqrt{\Delta t}, \\
\Delta^i b_t &= -\delta^b b_t \sqrt{\Delta t}.
\end{aligned}$$

Intuitively, in order to avoid having to fire-sale illiquid securities at a cost λ^s, deficit banks mobilize as much as they can from their other (more liquid) asset holdings. Note that all losses from a deficit bank are gained by a surplus bank. Therefore, assuming that $\sigma_t^d d_t > \delta^m m_t + \delta^p p_t + \delta^b b_t$, the law of motion of bank wealth in the illiquid stage can be written as

$$\Delta^i n_t = \lambda^i \Big(\sigma_t^d d_t - \theta^m m_t - \theta^p p_t - \theta^b b_t \Big) \varepsilon_t^i \sqrt{\Delta t}.$$

where $\theta^j \equiv \frac{\lambda^j - \lambda^s}{\lambda^s} \delta^j$ for $j \in \{m, p, b\}$ is defined as the liquidity index of a given asset, taking into account the liquidity frictions on prices and on quantities.

Let's now consider the case in which liquidity is sufficient to cover a negative funding shock: $\sigma_t^d d_t \leq \delta^m m_t + \delta^p p_t + \delta^b b_t$. In this case, the deficit bank does not have to pay any securities' fire-sale cost but still has to cover the cost of using liquid assets. Computing this cost requires knowing which assets have been used. Using a similar logic as previously, the deficit bank

[23] The problem is static since banks are able to fully readjust their balance sheets at the beginning of the next period.

will always first use less costly assets. In order to avoid dealing with multiple kinks and keep the model tractable in its continuous-time approximation, we make the following technical assumption.

Assumption 1 (Costless Liquidity Absent Fire-sale Risk). *When there is no fire-sale risk, $\sigma_t^d d_t \leq \delta^m m_t + \delta^p p_t + \delta^b b_t$, there is no cost of mobilizing liquid assets $\lambda^m = \lambda^b = \lambda^p = 0$.*

When Assumption 1 holds, the threshold at which banks do not have to fire sale securities corresponds to the threshold at which liquidity risk is nil and the law of motion for the wealth of banks is given by

$$\Delta^i n_t = 0.$$

Continuous-time Approximation We can combine the law of motion of both stages to get

$$\begin{aligned}\Delta n_t &= \Delta^\ell n_t + \Delta^i n_t \\ &= \Big(r_t^i i_t + r_t^m m_t + r_t^p p_t + r_t^b b_t - r_t^d d_t - c_t n_t + \mu_t^\tau n_t\Big)\Delta t \\ &+ \lambda^s \max\left\{\sigma_t^d d_t - \theta^m m_t - \theta^p p_t - \theta^b b_t, 0\right\}\varepsilon_t^i \sqrt{\Delta t}.\end{aligned}$$

Finally, the limit when Δt tends to 0 is given by

$$\begin{aligned}dn_t &= \Big(r_t^i i_t + r_t^m m_t + r_t^p p_t + r_t^b b_t - r_t^d d_t - c_t n_t + \mu_t^\tau n_t\Big)dt \\ &+ \lambda^s \max\left\{\sigma_t^d d_t - \theta^m m_t - \theta^p p_t - \theta^b b_t, 0\right\}d\widetilde{Z}_t,\end{aligned}$$

where $\widetilde{Z}_t$ is an idiosyncratic Brownian motion.