

NO. 1019
MAY 2022

REVISED
NOVEMBER 2025

Scarce, Abundant, or Ample? A Time-Varying Model of the Reserve Demand Curve

Gara Afonso | Domenico Giannone | Gabriele La Spada |
John C. Williams

Scarce, Abundant, or Ample? A Time-Varying Model of the Reserve Demand Curve

Gara Afonso, Domenico Giannone, Gabriele La Spada, and John C. Williams

Federal Reserve Bank of New York Staff Reports, no. 1019

May 2022; revised November 2025

JEL classification: E41, E43, E52, E58, G21

Abstract

What level of central bank reserves satiates banks' demand for liquidity? We estimate the slope of the reserve demand curve in the U.S. over 2010–2024 using a time-varying instrumental-variable approach at the daily frequency. When reserves exceed 12-13 percent of banks' assets, demand for reserves is satiated and reserves are abundant; below this threshold, the curve's slope becomes increasingly negative as reserves decline from ample to scarce. We also find that reserve demand has shifted over time, both vertically and horizontally, and identify important drivers of vertical shifts. Our methodology works well out-of-sample and can assess reserve ampleness in real time.

Key words: demand for reserves, federal funds market, monetary policy

Afonso, La Spada, Williams: Federal Reserve Bank of New York (emails: gara.afonso@ny.frb.org, gabriele.laspada@ny.frb.org, john.c.williams@ny.frb.org). Giannone: University of Washington, CEPR (email: dgiannon@uw.edu). The authors thank Roc Armenter, Marco Cipriani, Richard Crump, Marco Del Negro, Darrell Duffie, Thomas Eisenbach, Huberto Ennis (discussant), Antoine Martin, Enrico Perotti (discussant), A. Lee Smith (discussant), Andrea Tambalotti, Quentin Vandeweyer (discussant), Miklos Vari (discussant), and participants at the 2021 Banque de France Conference on Real-Time Data Analysis, Methods and Applications, the 2021 ECB Conference on Money Markets, the 2022 SNB Research Conference, FIRS 2023, the 2023 Dolomiti Macro Meetings, SAET 2023, the 2023 System Conference On Financial Institutions, Regulation, and Markets, the 2024 BCBS-BIS-CEPR conference on Banks' Liquidity in Volatile Macroeconomic and Market Environments, the 2024 LSE Adam Smith Workshop, the 2024 WFA, and at numerous seminars for valuable feedback. They are grateful to Logan Casey, JC Martinez, Peter Prastakos, Eric Qian, Nish Sinha, and Mihir Trivedi for excellent research assistance.

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the author(s) and do not necessarily reflect the position of the Federal Reserve Bank of New York, the Federal Open Market Committee, or the Federal Reserve System. Any errors or omissions are the responsibility of the author(s).

To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr1019.html.

1 Introduction

What level of central bank reserves satiates banks’ demand for liquidity? We answer this question by studying the demand curve for reserves in the U.S. banking system. This curve describes the price at which banks are willing to trade their reserve balances—the federal funds rate—as a function of aggregate reserves in the system. Economic theory predicts that banks’ demand for reserves is “satiated” above a given reserve level. Above this satiation point, reserves are *abundant*: the demand curve is flat, and the federal funds rate does not respond to changes in the aggregate supply of reserves. Below the satiation point, the relationship between price and quantity becomes increasingly negative, as reserves decline from *ample*—where the demand curve is gently sloped—to *scarce*—where the curve is steeply sloped. The large swings in the supply of reserves caused by the expansions and contractions of the Federal Reserve’s balance sheet over the last 15 years provide a rich data set to test this theory and estimate, for the first time, the satiation point in the demand for reserves.

Knowing the slope of the reserve demand curve at different reserve levels and its satiation point is important to implement monetary policy successfully and prevent money market instabilities. Since 2008, the Federal Open Market Committee (FOMC), as many other central banks, has operated a floor system. In such a system, the supply of reserves is sufficiently large that control over short-term rates is exercised primarily through changes in administered rates such as the interest rate on reserve balances (IORB), and active management of the reserve supply is not required (FOMC, 2019). Implementing a floor system thus requires knowing the level of reserves at which rates stop responding to reserve shocks; i.e., the minimum level of reserves that satiates banks’ demand for them. This knowledge helps avoid operating near or inside the steeply-sloped region of the demand curve, where modest reserve shocks cause material price changes. Although reserves have fluctuated between one and four trillion dollars since 2008, recent episodes of money market dislocations, as September 2019 and March 2020, suggest that, at times, reserves were scarce relative to the needs of the banking system (Afonso et al., 2021; d’Avernas et al., 2024).

In this paper, we develop a method to estimate the elasticity of the federal funds rate to reserve shocks (the slope of the reserve demand curve) at a daily frequency and identify the point of demand satiation, where the curve transitions between the flat and gently-sloped regions. We provide the first structural estimates of the different slopes of the demand curve for reserve levels ranging from scarce to abundant, using 15 years of data, from 2010 to 2024.

Estimation of the slope of the reserve demand curve is challenging for three reasons. First, economic theory predicts that the reserve demand curve is highly nonlinear, transi-

tioning from steeply downward-sloping to flat as the supply of reserves increases. Second, the demand for reserves moves over time due to structural changes in banking regulation and supervision, banks' internal risk management, and market functioning. Third, estimation of the demand for reserves is subject to endogeneity issues due to confounding factors in the demand equation. For example, the FOMC responds to unusual dislocations in money markets by increasing reserves, as it did in September 2019 and March 2020. Moreover, aggregate reserves also change due to factors that are outside the Federal Reserve's control and that are correlated with banks' demand for liquidity, such as the U.S. Treasury's account with the Federal Reserve and the overnight reverse repurchase agreement facility.

Our estimation strategy addresses these three challenges. Instead of estimating a non-linear function with possible slow-moving structural shifts, we estimate the slope of a linear function with time-varying coefficients and stochastic volatility at the daily frequency. That is, on each day, we estimate the slope of the tangent line to the reserve demand curve at the level of reserves attained on that day. With this locally linear approach, we trace the non-linear shape of the curve over time by moving along the curve, while allowing for low-frequency movements of the curve. Our specification is agnostic about the economic forces moving the curve over time, which allows for very general types of structural changes; subject to the assumption that the parameters of the time-varying linear model evolve more slowly than the daily liquidity shocks affecting banks' demand for reserves.

We address potential endogeneity issues by using a time-varying instrumental-variable (IV) approach. We instrument reserves in our linear approximation with past forecast errors from a daily time-varying vector autoregressive (VAR) model of the joint dynamics of reserves and federal funds rates, based on Primiceri (2005) and Del Negro and Primiceri (2015). The idea underlying our identification strategy is that in the current monetary policy framework, the Federal Reserve responds to fluctuations in the federal funds rate quickly but not on the same day; as a result, exogenous supply shocks can be extracted by a daily reduced-form forecasting model of reserve supply that captures the bulk of reserve variation in our sample (i.e., quantitative easing and tightening). Importantly, using daily data allows us to control for the FOMC response to rate dislocations and for short-term disturbances to demand related to factors outside the Federal Reserve control.

Our time-varying estimates of the rate elasticity to reserve shocks imply that the satiation point in the reserve demand curve, where reserves transition from abundant to ample, occurs for reserves around 12-13% of bank assets, depending on the period. The curve becomes increasingly steep as reserves decline, with reserves approaching scarcity around 8-10% of bank assets. Furthermore, we obtain qualitatively similar results if we normalize reserves

by bank deposits or GDP, confirming that the choice of the normalization factor does not materially alter the shape of the curve but simply removes a time trend in nominal reserves.

Our results are robust to controlling for changing conditions in repo and Treasury markets. In addition, we replicate our analysis using a structural time-varying VAR based on sign restrictions (Uhlig, 2005; Rubio-Ramírez et al., 2010) to extract the supply shocks used as instruments in our IV methodology and find similar results to our baseline specification.

To validate our results, we also compare our estimates of the reserve demand elasticity over time with other measures of reserve amplexness, which are not based on fluctuations in the federal funds rate: the share of overnight Treasury repo with rates above the IORB, the share of late payments in the Fedwire Funds Service system, the share of domestic federal funds borrowing, and the banking industry average intraday overdraft. All indicators co-move with our elasticity estimates, supporting the validity of our methodology and findings.

In addition, we show that because our forecasting model of reserve supply has good predictive accuracy out-of-sample, our methodology can be used in real time to monitor the relative amplexness of the supply of reserves. For example, if used in real time during the Federal Reserve’s balance-sheet normalization of 2015–2019, our methodology would have indicated that we were entering the negatively sloped region of the reserve demand curve six to nine months ahead of the events of September 2019. The real-time applicability of our methodology is particularly useful for the implementation of quantitative tightening and the assessment of its implications.

Finally, we provide evidence of both vertical and horizontal low-frequency shifts in the demand for reserves during 2010–2024, with upward vertical shifts being particularly material. Using our estimates of the slope of the reserve demand curve, we are able to identify two drivers of structural vertical shifts: the spread between the Overnight Reverse Repo (ONRRP) rate and the IORB, and bank balance-sheet costs. An increase in the ONRRP rate relative to the IORB increases the reservation price of Federal Home Loan Banks (FHLBs) when lending in the federal funds market, pushing the federal funds rate up; an increase in balance-sheet costs increases banks’ costs of borrowing, pushing the rate down.

The idea behind our identification of vertical shifts is that in the flat region of the reserve demand curve, neither supply shocks nor horizontal demand shifts can move the federal funds rate: persistent changes in the rate can only be due to structural vertical shifts in the demand for reserves. We then study the response of the federal funds rate to changes in the ONRRP-IORB spread and in bank balance-sheet costs that were implemented in periods when our estimates of the reserve demand elasticity are insignificant (i.e., zero slope).

We find that the ONRRP-IORB spread adjustments of 2014–2015 had a pass-through to the federal funds-IORB spread of around 40% on the day of the adjustment and of 60–100% after one month; this evidence is consistent with an upward shift in the reserve demand curve in the second half of our sample due to the introduction of the ONRRP facility in late 2013 and to the overall increase in the ONRRP rate relative to the IORB since then. We also find that the Supplementary Leverage Ratio (SLR) relief of 2020–2021, by temporarily decreasing banks’ balance-sheet costs, pushed the federal funds-IORB spread up by 3 basis points (bp) after one month from the relief introduction and by 7 bp after two months; this evidence is consistent with a significant upward shift in the reserve demand curve during 2020–2021 (see Figure 2).

An implication of vertical shifts is that the level of the spread between the federal funds rate and the rate paid on reserves does not provide a sufficient statistic for the degree of ampleness of reserves, as is often assumed. In contrast to other approaches that rely exclusively on the level of this spread, our methodology focuses on the elasticity of rates to reserve shocks, which is more closely related to the policy goal of interest rate control.

This paper contributes to the extensive literature on the demand for reserves and its implications for monetary policy implementation and transmission to the economy. Among the early literature, Hamilton (1996, 1997) was the first to emphasize the importance of using daily data to identify the slope of the reserve demand curve. This early literature, however, studied periods of reserve scarcity and is not informative on the regions of ample and abundant reserves (Bernanke and Blinder, 1992; Christiano and Eichenbaum, 1992; Bernanke and Mihov, 1998a,b). We provide the first empirical counterpart to the new theoretical literature that focuses on the period after 2008 (Afonso et al., 2019; Bigio and Sannikov, 2021; Bianchi and Bigio, 2022; Afonso et al., 2023; Lagos and Navarro, 2023).

Recent empirical studies on the demand for reserves include Smith (2019), Smith and Valcarcel (2023), and Lopez-Salido and Vissing-Jorgensen (2023). Relative to these papers, we make four contributions: (1) we identify the satiation point in banks’ demand for reserves, (2) we provide structural time-varying estimates of the slope of the reserve demand curve using a novel identification scheme based on daily data, which allows us to control for both the endogeneity due to the FOMC’s actions and the endogeneity due to confounding factors that are outside the Federal Reserve’s control, (3) we document low-frequency shifts in the demand curve, and (4) we show that our methodology can be used in real time as a monitoring tool to detect tightness in the market for reserves.

Finally, our paper provides independent empirical evidence related to the growing literature on central bank reserves and money-market dislocations (Correa et al., 2020; Afonso

et al., 2021; Copeland et al., 2025; d’Avernas et al., 2024; d’Avernas and Vandeweyer, 2024) and the implications of central bank balance-sheet expansions and contractions (Stein, 2012; Kashyap and Stein, 2012; Greenwood et al., 2016; Diamond et al., 2024; Acharya and Rajan, 2024; Benigno and Benigno, 2022; Acharya et al., 2023).

The remainder of the paper is organized as follows. Section 2 presents a model of the demand for reserves and discusses the institutional setting and endogeneity issues. Section 3 describes the data used in the estimation. Section 4 discusses the econometric model and the identification strategy. Section 5 reports the results of our time-varying IV estimation of the reserve demand elasticity. Section 6 identifies two drivers of vertical shifts in the reserve demand curve. Section 7 concludes. The appendix includes an extension of the model with bank balance-sheet costs and a detailed description of the forecasting model. The internet appendix reports a formal out-of-sample validation of the forecasting model, robustness checks of the elasticity estimates, additional evidence on the drivers of vertical shifts, and results of a nonlinear fit of the reserve demand curve.

2 The Demand for Reserves

2.1 Theory

We present a simple model of the demand for reserves that builds on the seminal paper by Poole (1968).¹ There are N risk-neutral banks, and two periods: intraday and end-of-day. Banks hold initial reserves \tilde{r}_i in their accounts at the central bank and target an end-of-day reserve level \bar{r}_i to process payments and meet regulatory and internal liquidity requirements. End-of-day reserve balances pay an overnight interest rate i^{IORB} set by the central bank. At the end of the day, banks can also borrow reserves from the central bank’s discount window (DW) at a rate $i^{DW} > i^{IORB}$.²

During the day, banks monitor their balances and borrow ($f_i > 0$) or lend ($f_i < 0$) reserves in a competitive (federal funds) market at a rate i^f to meet their end-of-day target levels.³ After the federal funds market closes, however, banks are subject to a late-day liquidity shock z_i that changes their reserve balances—e.g., an unexpected late payment or delayed accounting information. As a result, bank i enters the end-of-day period with a

¹More recent papers include Ennis and Keister (2008), Afonso and Lagos (2015), Armenter and Lester (2017), Afonso et al. (2019), Bigio and Sannikov (2021), Bianchi and Bigio (2022), Lagos and Navarro (2023).

²We can also think of i^{DW} as the minimum bid rate at the Standing Repo Facility (SRF), another standing facility offered by the Federal Reserve; the DW and SRF rates are currently set at the same level.

³A federal funds transaction is an unsecured loan in U.S. dollars, typically of overnight maturity.

stochastic balance equal to $r'_i \equiv \tilde{r}_i + f_i + z_i$, where z_i is distributed according to a cumulative distribution function G with support $(-\underline{z}, \bar{z})$. For simplicity, we assume that G is absolutely continuous, with density g , and the same for every bank.

At the end of the day, banks compare their end-of-day balances with their targets. If a bank ends the day with a deficit ($r'_i < \bar{r}_i$), it borrows $\bar{r}_i - r'_i$ from the central bank at i^{DW} .

When a bank decides how much to borrow or lend in the federal funds market, it takes into account the uncertainty introduced by the late-day shock z_i . Namely, the bank chooses the optimal federal funds borrowing f_i that minimizes the expected opportunity cost of holding end-of-day reserve balances, taking all rates as given. That is,

$$\min_{f_i} \left[(i^f - i^{IORB}) \int_{\hat{z}_i}^{\bar{z}} (r'_i - \bar{r}_i) g(z) dz + (i^{DW} - i^f) \int_{\underline{z}}^{\hat{z}_i} (\bar{r}_i - r'_i) g(z) dz \right], \quad (1)$$

where $\hat{z}_i \equiv \bar{r}_i - \tilde{r}_i - f_i$. The first term in (1) captures the cost of ending the day with excess balances ($r'_i - \bar{r}_i > 0$), earning interest on those reserves, but forgoing the opportunity to lend (or to have borrowed less) at a rate i^f in the market. The second term captures the cost of ending the day with reserves below the target, borrowing $\bar{r}_i - r'_i$ from the central bank at the rate i^{DW} instead of borrowing in the market during the day at a rate i^f .

The first order condition yields bank i 's optimal federal funds borrowing f_i^* :

$$i^f = i^{IORB} + (i^{DW} - i^{IORB}) G(\bar{r}_i - (\tilde{r}_i + f_i^*)), \quad (2)$$

where f_i^* is the unique solution that minimizes equation (1).⁴

Equation (2) characterizes bank i 's inverse demand for reserves $i^f(r_i)$, where $r_i = \tilde{r}_i + f_i^*$ is the bank's total reserves after the market closes and before the liquidity shock is realized. Since G is a cumulative distribution function, the demand curve is decreasing and bounded between i^{DW} and i^{IORB} . If the distribution of liquidity shocks G has unbounded support, the demand is strictly decreasing, converging to i^{DW} as reserves decrease and to i^{IORB} as reserves increase. If G has bounded support, the curve is strictly decreasing on $[\bar{r}_i - \bar{z}, \bar{r}_i + \underline{z}]$ and flat outside this interval, with $i^f(r_i) = i^{DW}$ for $r_i \leq \bar{r}_i - \bar{z}$ and $i^f(r_i) = i^{IORB}$ for $r_i \geq \bar{r}_i + \underline{z}$.

We can derive the aggregate demand for reserves by inverting equation (2) to obtain bank i 's demand for reserves as a function of the federal funds rate, $f_i^*(i^f)$, and then

⁴It is straightforward to show that optimization problem (1) has a unique minimum by taking the second derivative of the objective function, which is non-negative everywhere and strictly positive on G 's support.

summing reserves across banks to obtain the aggregate demand $r \equiv \sum_{i=1}^N r_i = \sum_{i=1}^N \bar{r}_i - NG^{-1} \left(\frac{i^f - i^{IORB}}{i^{DW} - i^{IORB}} \right) \equiv \bar{r} - NG^{-1} \left(\frac{i^f - i^{IORB}}{i^{DW} - i^{IORB}} \right)$, where $\sum_{i=1}^N f_i^* = 0$.⁵ Inverting this expression back, we obtain the aggregate (inverse) demand curve for reserves:

$$i^f = i^{IORB} + (i^{DW} - i^{IORB})G \left(\frac{\bar{r} - r}{N} \right). \quad (3)$$

The demand for reserves in equation (3) is a nonlinear, decreasing function with an upper asymptote equal to i^{DW} (as reserves decrease) and a lower asymptote equal to i^{IORB} (as reserves increase). If the distribution of liquidity shocks has bounded support, the curve becomes perfectly flat at $i^f = i^{IORB}$ ($i^f = i^{DW}$) if reserves are sufficiently large (small).⁶

An intuitive way to interpret the demand curve (3) is in terms of arbitrage conditions. If the federal funds rate i^f were higher than the discount rate i^{DW} , no bank would be willing to borrow in the federal funds market: banks would borrow arbitrarily large amounts from the central bank and lend their reserves in the federal funds market, driving their demand to zero and earning the spread between the federal funds and discount-window rates (DW arbitrage). Similarly, if i^f were lower than the interest paid on reserves i^{IORB} , banks would borrow arbitrarily large amounts in the federal funds market and hold these reserve balances at the central bank, pushing their demand for reserves to infinity and earning the spread between the IORB and federal funds rates (IORB arbitrage). Between the two asymptotes, the relationship between prices and quantities is negative: as the federal funds rate decreases, the opportunity cost of holding reserves declines, and the relative cost of borrowing from the central bank raises, increasing banks' demand for reserves (Swanson, 2023).

Equation (3) is derived from a model without frictions. In practice, however, there are frictions in the federal funds market that affect the reserve demand curve. An example is market segmentation. Federal Home Loan Banks (FHLBs)—a type of government-sponsored enterprise—are the main lenders in the federal funds market but do not earn the IORB rate on their balances with the Federal Reserve; they can, however, invest at the Federal Reserve through the Overnight Reverse Repo facility (ONRRP), earning a fixed rate below the IORB rate.⁷ FHLBs are therefore willing to lend at rates below the IORB rate but above the ONRRP rate. As a result, the lower bound of the reserve demand curve lies between the

⁵If G has bounded support, aggregate demand is $r \leq \sum_{i=1}^N (\bar{r}_i - \bar{z}) = \bar{r} - N\bar{z}$ for $i^f = i^{DW}$, $r = \bar{r} - NG^{-1} \left(\frac{i^f - i^{IORB}}{i^{DW} - i^{IORB}} \right)$ for $i^{DW} > i^f > i^{IORB}$, and $r \geq \sum_{i=1}^N \bar{r}_i + \bar{z} = \bar{r} + N\bar{z}$ for $i^f = i^{IORB}$.

⁶Namely, $i^f = i^{IORB}$ for $r \geq \bar{r} + \bar{z}$, and $i^f = i^{DW}$ for $r \leq \bar{r} - \bar{z}$.

⁷The Federal Reserve adjusts the ONRRP and IORB rates in the implementation of monetary policy. See Cipriani and La Spada (2022) and Afonso et al. (2022a) for details.

IORB and ONRRP rates, with the distance from these bounds depending on the bargaining power of the FHLBs relative to that of the borrowing banks.

Another important friction affecting the reserve demand curve is banks’ balance-sheet costs; i.e., all those costs—either regulatory or internal—that are based on balance-sheet size. These costs penalize borrowing, especially to finance safe assets such as reserves, limiting banks’ ability to take arbitrage positions (Duffie, 2018; Anderson et al., 2020). Notable examples are the Federal Deposit Insurance Corporation (FDIC) assessment fee and the Basel III leverage-ratio regulation. As a result of these costs, when borrowing to conduct the IORB arbitrage, banks demand a rate lower than the IORB rate, pushing the lower asymptote of the reserve demand curve below i^{IORB} . Similarly, the upper asymptote will be above i^{DW} : when lending to engage in DW arbitrage, banks demand a higher rate than the rate at which they borrow from the central bank.⁸

In Appendix A, we present two model extensions that incorporate FHLBs and bank balance-sheet costs.⁹ In particular, in addition to deriving the reserve demand curve in the presence of these frictions, we show that the vertical location of the curve’s lower asymptote (i.e., the level of the federal funds rate at which bank’s demand is satiated) increases with the ONRRP rate—the FHLBs’ outside option—and with their bargaining power, while it decreases with banks’ balance-sheet costs. We test these theoretical predictions in Section 6.

Figure 1 provides an illustrative example of the reserve demand curve in the presence of market frictions. Over the past decade, the federal funds rate has mostly remained close to the IORB rate and consistently below the DW rate, suggesting that the banking system has operated away from the curve’s upper asymptote. For this reason, from now on, we focus on the right part of the curve, which can be divided into three regions (see Figure 1): a region of scarce reserves, where the curve is steep, a region of abundant reserves, where the curve is flat, and an intermediate region of ample reserves, where the curve is gently sloped.¹⁰

This section yields five important takeaways about the reserve demand curve. (1) It is convenient to express the price of reserves as the federal funds rate minus the IORB rate to control for changes in the opportunity cost of lending reserves determined by the

⁸Another friction affecting the curve’s upper asymptote and pushing i^f above i^{DW} is the stigma associated with borrowing from the discount window (Armantier et al., 2015).

⁹See also Afonso et al. (2019) and Kim et al. (2020) for theoretical models incorporating these frictions.

¹⁰The model in d’Avernas et al. (2024) shows that, in addition to shifting the curve to the right, the presence of intraday liquidity requirements introduces a vertical asymptote. While our model does not capture this feature, it allows for the slope to approach minus infinity (e.g., if the distribution of liquidity shocks is highly concentrated). Importantly, as we explain in Section 4, our empirical strategy is robust to the presence of a vertical asymptote as long as the asymptote is not in the flat region of the demand curve (i.e., the curve is not a step function) and the supply shocks in our IV estimation are sufficiently small.

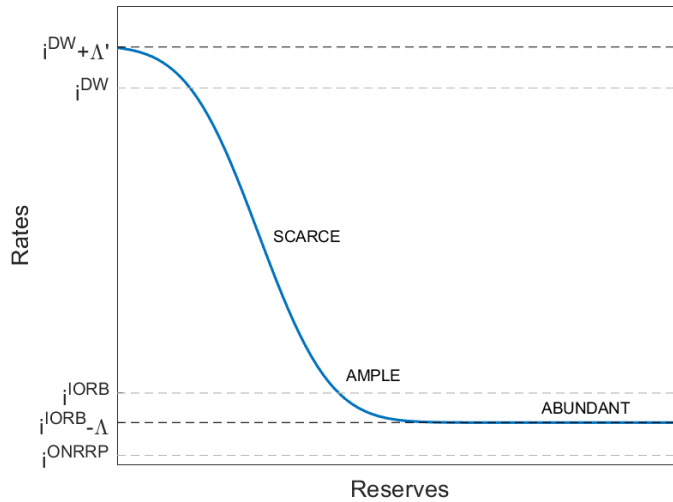


Figure 1: **Bank demand for reserves in the presence of market frictions.** Example of reserve demand curve with banks’ balance-sheet costs, market segmentation, and discount-window stigma. i^{DW} , i^{IORB} , and i^{ONRRP} represent the DW, IORB, and ONRRP rates. Λ captures the effects of balance-sheet costs and market segmentation on the lower asymptote; Λ' captures the effects of balance-sheet costs and stigma on the upper asymptote.

FOMC monetary policy stance. (2) Aggregate reserves should be normalized by a measure of the size of the banking system (N in the simple model above). (3) The curve is highly nonlinear and flattens as reserves increase. (4) Changes in banks’ balance-sheet costs or in FHLBs’ relative outside option (ONRRP rate relative to IORB rate) vertically shift the lower asymptote of the curve. (5) The horizontal location of the curve is affected by banks’ aggregate reserve target (\bar{r}), which itself can move over time due to changes in liquidity regulation, risk management, and market structure.

2.2 Monetary policy implementation

Knowing the slope of the reserve demand curve at different reserve levels is key to monetary policy because the FOMC uses the federal funds rate to communicate the policy stance. If reserves are scarce, for instance, supply shocks move this rate significantly, whereas they have no effect if reserves are abundant. As many other central banks after 2008, the Federal Reserve operates a system in which the supply of reserves is sufficiently large that typical reserve fluctuations do not move prices significantly, and rate control is implemented through changes in the IORB and other administered rates (FOMC, 2019).

Operating in the flat or almost-flat portion of the reserve demand curve (“floor system”) has several advantages relative to implementing monetary policy in the steeply-sloped region (“corridor system”). (1) It improves rate control because the response of rates to supply shocks is minimal (or zero if reserves are abundant). (2) For the same reason, it has lower operational costs, as it requires less active reserve management and less frequent operations in response to shocks. (3) By divorcing the amount of liquidity from the level of rates, the central bank can use reserves as a separate tool to address financial stability issues, while using administered rates to implement the monetary policy stance (Keister et al., 2008). (4) Higher reserve levels reduce the risk of disruptions in gross-settlement payment systems, like Fedwire Funds Services in the U.S. or T2 in the Eurozone, because banks rely less on intraday credit and incoming payments to process outgoing payments (Afonso et al., 2022b; Copeland et al., 2025; Ding et al., 2025). (5) Finally, by providing an ample supply of safe short-term claims, the central bank can weaken financial intermediaries’ incentives to issue runnable short-term liabilities (Hanson et al., 2015; Greenwood et al., 2016).

Of course, floor frameworks also have drawbacks. (1) A large balance sheet exposes the central bank to political economy costs, including concerns about fluctuations in its income and about the weight of government debt in its balance sheet (Bindseil, 2016). (2) Abundant reserves discourage unsecured interbank trading and active reserve redistribution, hindering market discipline and price discovery (Borio, 2023). (3) By expanding banks’ balance sheets with a low-return asset, reserve injections tighten bank balance-sheet costs, leading banks to reduce low-margin intermediation and lending (Duffie, 2018; Diamond et al., 2024). (4) Finally, reserve injections may be associated with an increase in banks’ uninsured demand deposits and credit lines; if these claims on liquidity do not shrink during quantitative tightening, banks may become more vulnerable to shocks (Acharya and Rajan, 2024; Acharya et al., 2023).

Based on these benefits and costs, we define two main types of inadequate estimation of the slope of the reserve demand curve in a floor system depending on whether the point of demand satiation is underestimated (i.e., the curve is steep at higher reserve levels than expected) or overestimated (i.e., the curve is steep at lower levels than expected). In the first case, the central bank may operate too far to the left of the satiation level; in the second one, it may operate too far to the right.¹¹

Quantifying the costs associated with these estimation errors is challenging because they include reputational and political economy considerations and are often defined relative

¹¹See Afonso et al. (2023) for a formal treatment of the central bank choice of optimal reserve supply with a nonlinear reserve demand curve as in equation (3) and uncertainty about its location and steepness.

to socially efficient outcomes that are not readily measurable. One quantifiable cost of underestimating the point of demand satiation is the cost of money-market volatility. A notable example is the turmoil of September 2019, which was likely due to reserve scarcity (Afonso et al., 2021; d’Avernas et al., 2024; Copeland et al., 2025). According to our time-varying estimates of the slope of the reserve demand curve, in mid-September 2019, reserves were below the most recent estimate of the satiation point (early 2018) by 5 percentage points (pp) of commercial bank assets (Section 5). Consistent with this result, around a moderate reduction in reserve supply, the annualized (volume-weighted) average federal funds rate jumped to 2.39 pp on September 17, from 2.14 pp on September 13, an unprecedented increase of 12% that led the policy rate to breach the top of its target range.

A quantifiable cost of overestimating the point of demand satiation is the cost of crowding out bank lending. Diamond et al. (2024) find that for each dollar of reserves injected in 2008–2017, bank lending decreased by 7 cents. Based on these estimates, overestimating the satiation level by 5 pp of bank assets in September 2019—an error of similar magnitude as in the scenario described above—would have reduced bank lending by \$70 billion.

Importantly, the costs associated with an excessive supply of reserves indicate that the issue of adequately estimating the satiation level in the reserve demand curve cannot be bypassed by simply providing a surplus of reserves above a rough estimate.

2.3 Reserve supply and endogeneity

To understand the sources of endogeneity in the demand for reserves, it is important to understand how reserves change. Reserves are balances that banks hold in their accounts at the Federal Reserve and change for two reasons: either because the Federal Reserve buys or sells securities, which happens through the banking system, or because funds are transferred between reserves and non-reserve accounts at the Federal Reserve. Some institutions, such as money market funds or the Department of the Treasury, have Federal Reserve accounts and transact with banks; when these transfers occur, reserves change outside the Federal Reserve’s control.¹² Both types of reserve fluctuation can be endogenous to bank demand.

The first type of endogeneity is due to the Federal Reserve’s actions. On occasion, the Federal Reserve responds to unusual volatility in the federal funds rate by adjusting the supply of reserves to keep the rate within its target range (e.g., September 2019, March 2020).¹³ These responses are quick and put in place within a matter of days; in September

¹²In contrast, interbank transactions redistribute reserves across banks, without affecting the aggregate.

¹³Before 2008, when reserves were very scarce, this type of endogeneity was more severe because the Federal

2019, for instance, the Federal Reserve started to expand the supply of reserves on the 18th, and rates returned to their prior levels shortly afterwards (Afonso et al., 2021). If demand shocks are serially correlated, the Federal Reserve’s response today—a function of yesterday’s demand shock—will be correlated with today’s demand shock.

The second type of endogeneity is due to the Federal Reserve’s non-reserve liabilities that are correlated with banks’ demand for funding. An example is the Treasury General Account (TGA) with the Federal Reserve. When banks buy Treasuries at auction, they transfer funds from their Federal Reserve accounts to the TGA, which results in an increase in the TGA and a decrease in reserves. Around the same time, banks’ demand for short-term funding increases because they finance these purchases with overnight repos. The temporary increase in repo rates can put upward pressure on the federal funds rate because federal funds and repos are close substitutes (Schulhofer-Wohl and Clouse, 2018).

Martin et al. (2019) argue that the issuance of Treasuries also affects the federal funds rate by placing upward pressure on Treasury yields. When Treasury yields go up, money market funds (MMFs), which hold a large share of Treasuries in their portfolios, become a more attractive investment than bank deposits. This competitive pressure induces banks to increase their overnight wholesale deposit rates, such as the Eurodollar rate. Since overnight wholesale deposits and federal funds are substitute forms of funding, the surge in deposit rates can push the federal funds rate up.¹⁴

Another example of non-reserve liability correlated with the federal funds rate is the ONRRP. Through this facility, both FHLBs and MMFs—the main lenders in the federal funds and repo markets, respectively—can invest at the Federal Reserve via fixed-rate overnight repos. Variations in the ONRRP balance correlate with the demand for reserves through the “window dressing” of European banks around reporting dates. While U.S. banks report daily averages of their leverage ratio on a quarterly basis, the Basel III leverage ratio for European banks is calculated using only quarter-end or month-end data, depending on the reporting period. This reporting requirement gives European banks an incentive to temporarily reduce their overnight borrowing around those dates (Banegas and Tase, 2020, Bassi et al., 2024). This window dressing has two effects on the market for reserves. First, it lowers rates as the demand for federal funds declines. Second, it lowers reserves because MMFs invest more in

Reserve regularly adjusted the supply of reserves through daily open market operations to control the federal funds rate. This endogeneity has disappeared since the Federal Reserve has adopted a floor system.

¹⁴Another confounding factor related to the TGA are tax payments. When MMF investors use their fund shares to pay taxes, the fund instructs its custodian bank to submit the payment to the Treasury, resulting in a decline in reserves and an increase in the TGA. Around the same time, MMFs reduce their overnight lending to meet redemptions, pushing money-market rates up. Both Treasury settlements and tax payments played important roles in the turmoil of September 2019 (Afonso et al., 2021).

the ONRRP to offset the reduction in funding demand; since ONRRP trades settle on the books of a clearing bank, the result is an equal decrease in bank reserves.

More generally, conditions in the repo and Treasury markets not only affect bank demand for reserves by influencing the price of substitute forms of overnight funding, but are also correlated with reserve supply via changes in the TGA and ONRRP due to Treasury issuance and MMF investment. This type of endogeneity has become more important over time. First, Treasury repos have become more relevant for both bank funding and MMF lending due to the 2014 MMF reform (Anderson et al., 2020); (Cipriani and La Spada, 2021). Second, the variability of the TGA and ONRRP has increased (Afonso et al., 2020): excluding currency in circulation, they amounted to 5% of reserves in 2010 and 35% in 2024, with a peak of 79% in 2022 (Figure IA.1 in Internet Appendix A).

In Section 4, we propose a general instrumental-variable methodology to control for both the endogeneity due to the Federal Reserve’s actions and that due to non-reserve liabilities.

3 Data

3.1 Data description

We use daily data from January 1, 2009 to December 13, 2024. Since the federal funds market is only open on business days, we drop weekends and federal holidays, including Mondays following holidays that fall on a Sunday.¹⁵

Using daily data is key for our identification strategy because it allows us to directly address the endogeneity issues due to the “window dressing” of European banks around reporting dates. The regulation-induced reduction in European banks’ short-term borrowing reverts within a day or two. To control for this high-frequency and transitory omitted variable in the reserve demand, we exclude one-day windows around month-ends.

Our variables of interest are aggregate reserves and the federal funds rate. To calculate reserves, we use confidential daily data on the aggregate balances of depository institutions from internal Federal Reserve accounting records.¹⁶ To take into account the growth of the banking system over time, we normalize reserves by commercial banks’ assets. Weekly data on bank assets are available from the Federal Reserve Economic Data, FRED

¹⁵We use the holiday schedule for the Federal Reserve System and keep business days according to the `isBusinessDay` function in the TIS package <https://cran.r-project.org/web/packages/tis/tis.pdf>.

¹⁶A weekly version is available in the H.4.1 report (“Reserve balances with Federal Reserve Banks”).

(“TLAACBW027SBOG”). In robustness checks, we normalize reserves by bank deposits or GDP. Weekly data on deposits and quarterly data on GDP are available from FRED (“DP-SACBW027SBOG” and “GDP”). We linearly interpolate these series to obtain a daily series.

For the federal funds rate, we use the daily volume-weighted average, based on transaction data collected by the Federal Reserve Bank of New York.¹⁷ The underlying transactions data are those used for the official calculation of the effective federal funds rate (EFFR). As explained in Section 2.1, we subtract the IORB rate from the federal funds rate. Daily data on the IORB are available from FRED (“IOER” and “IORB”).

In robustness checks, we control for repo rates. We use the volume-weighted average rate of overnight repos collateralized by Treasuries (with maturity up to 30 years) cleared through the Fixed Income Clearing Corporation (FICC) GCF Repo service, which is available on the Depository Trust & Clearing Corporation (DTCC) website.¹⁸ We choose these data because overnight Treasury repos are the largest segment of the repo market and the most closely related to overnight federal funds transactions. In additional robustness checks, we control for the daily market yields on Treasuries at one-year constant maturity (FRED “DGS1”).

To validate our results, we compare our time-varying estimates of the reserve demand elasticity with other indicators of reserve ampleness based on data other than the federal funds rate. Daily data on the timing and volume of interbank payments are from the Fedwire Funds Service. Daily data on banks’ intraday overdrafts are from the Federal Reserve internal records.¹⁹ Daily data on federal funds borrowing by U.S. banks are from the Federal Reserve Bank of New York website.²⁰ Daily transaction data on the volume and rate of overnight Treasury repos in FICC’s GCF Repo service are from the Office of Financial Research.

To identify drivers of vertical shifts in the demand curve, we study the effect of the changes that the Federal Reserve implemented to the spread between the rate at the ONRRP facility and the interest paid on reserves. Daily data in the ONRRP are available from FRED (“RRPONTSYAWARD”). In our analysis, we control for high-frequency variations in risk aversion, credit risk, and interest risk rate using the daily VIX and TED spread from FRED (“VIXCLS”, “TEDRATE”) and the daily MOVE index from Bloomberg.

¹⁷The average from March 2016 onward is available at <https://www.newyorkfed.org/markets/reference-rates/>.

¹⁸<https://www.dtcc.com/charts/dtcc-gcf-repo-index>.

¹⁹A version of these data at the maintenance-period frequency is available at https://www.federalreserve.gov/paymentsystems/psr_dlod.htm.

²⁰<https://www.newyorkfed.org/markets/reference-rates/>.

3.2 Preliminary evidence

Panel (a) of Figure 2 shows the time evolution of reserves normalized by bank assets. Reserves went through a full expansion-contraction cycle from 2010 to late 2019 and expanded again from early 2020 to late 2021, ranging between 8% (2010 and 2019) and 19% (2014 and 2021) of bank assets; they started to decline again in 2022, fluctuating around 14–15% of bank assets until the end of 2024. These movements reflect the Federal Reserve balance-sheet expansions in response to the 2008 and 2020 crises, as well as the balance-sheet normalizations of 2015–2019 and of 2022–2024. Panel (b) plots the daily average federal funds rate minus the IORB rate. By comparing panels (a) and (b), we can see a negative correlation between quantities (reserves) and prices (federal funds rates), which suggests that after removing month-end data from our daily sample, supply shocks tend to dominate demand shocks.

This intuition is confirmed by panel (c), which plots realized rates against realized reserves and can be seen as an approximate visualization of the reserve demand curve. It shows a nonlinear, downward-sloping relationship between prices and quantities that flattens when reserves are sufficiently large (above the demand satiation point). Moreover, panel (c) shows that this relationship has moved outward over time: the curve moved up and to the right after 2014 and further up after March 2020.

Figure 2 shows equilibrium realizations over time and cannot be interpreted causally. To identify the slope of the reserve demand curve, we use a structural time-varying methodology.

4 Estimation Strategy

4.1 Time-varying linear approximation of the reserve demand curve

We can write the reserve demand curve in equation (3) of Section 2.1 as

$$p_t = p_t^* + f(q_t - q_t^*; \theta), \quad (4)$$

where p and q are the price and quantity of reserves, p^* is the curve’s lower asymptote, $f(x; \theta)$ is a decreasing non-linear function parameterized by θ that goes to zero as x goes to infinity, and q^* is the horizontal location of the curve relative to a normalization point.

As discussed in Section 2.1, we express p as the spread between the federal funds and IORB rates to control for changes in banks’ opportunity cost of lending reserves caused by the monetary policy stance. This formulation is consistent with recent theoretical deriva-

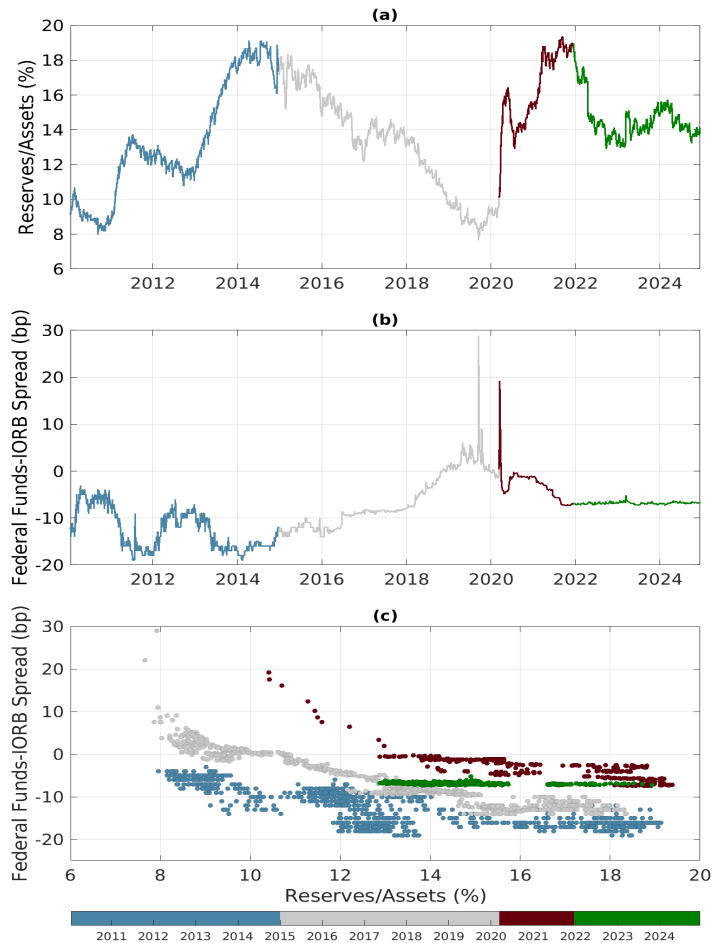


Figure 2: **Reserves, the federal funds rate, and the reserve demand curve.** Panel (a) plots reserves relative to bank assets from January 1, 2010 to December 13, 2024. Panel (b) shows the spread between the volume-weighted average federal funds rate and the IORB rate. Panel (c) plots the relationship between the spread and normalized reserves.

tions of the demand curve (Bigio and Sannikov, 2021; Bianchi and Bigio, 2022; Lagos and Navarro, 2023) and has been recently adopted by empirical papers (Lopez-Salido and Vissing-Jorgensen, 2023; Acharya et al., 2023). To account for the growth of the banking sector over time, as explained in Section 2.1, we measure q as reserves normalized by banks' total assets.

Variation in the curve's lower asymptote $p_t^* < 0$ arises from structural changes in banks' balance-sheet costs or in the outside option of FHLBs (see Section 2.1). Variation in the horizontal location of the curve q_t^* comes from structural changes in liquidity regulation, supervision, or market functioning that affect banks' demand for reserves.

Estimating the elasticity of the federal funds rate to reserve shocks is challenging for three reasons: (i) the demand curve (4) is a nonlinear function of reserves, which means that the slope is itself a function of aggregate reserves; (ii) persistent structural changes since the

2008 financial crisis may have moved the curve over time; and (iii) there are endogeneity issues due to the FOMC’s actions and to some Federal Reserve’s non-reserve liabilities.

To tackle the first two challenges, instead of estimating a nonlinear function with low-frequency shifts, we estimate the following linear model with time-varying coefficients at daily frequency:

$$p_t = \alpha_t + \beta_t q_t + \sigma_t v_t, \tag{5}$$

where p and q are the price and quantity of reserves as defined above, and v is a daily demand shock that can be serially correlated. All parameters, including the shock variance σ , can vary at daily frequency. β_t measures the elasticity of rates to reserves on each day. Model (5) is a locally linear approximation of the reserve demand curve (4) implied by the theory. Every day, we estimate the straight line that is tangent to the reserve demand curve at the level of reserves attained on that day; as reserve supply changes over time, these lines move and trace the curve. This approach enables us to capture the nonlinear nature of the curve, without specifying a functional form and allowing for low-frequency structural shifts.

Changes in the parameters of model (5) are due to either exogenous changes in the supply of reserves (i.e., movements along the curve) or structural changes in banks’ demand for reserves (i.e., persistent movements of the curve). The assumption behind model (5) is that its parameters evolve more slowly than the liquidity demand shocks that hit banks on a daily basis; this allows us to disentangle (low-frequency) variation in β from (high-frequency) variation in v . In our context, this assumption is plausible for two reasons. First, as we explain below, our estimation strategy uses small daily movements along the curve (so that the locally linear approximation works well) that reflect exogenous variation in expansions and contractions of the Federal Reserve balance sheet that took place over years.²¹ Second, the changes in regulation, supervision, and market structure that may have moved banks’ demand for reserves after 2008 were low-frequency events implemented over months. An important example of these regulatory changes is the SLR relief of 2020–2021, which temporarily shifted the curve upward. This policy was implemented during a period of large reserve injections by the Federal Reserve, raising potential endogeneity concerns. However, as we show in Section 6.2, it took several months for the banking system to fully adjust to this regulatory change. Consequently, the low-frequency vertical shift implied by the SLR relief is captured by the slow-moving intercept, α_t , while our estimation strategy identifies the slope of the demand curve, β_t , from high-frequency (daily) fluctuations in prices

²¹Strictly speaking, for the linear approximation, we also require the function f to be smooth, which is likely to hold because f represents the reverse cumulative distribution function of banks’ daily liquidity shocks (see Section 2.1).

and reserves.

4.2 Reserve supply and instrumental variable approach

To control for endogeneity, we use an instrumental variable (IV) approach. We propose a forecasting model of the joint dynamics of the quantity and price of reserves and use past forecast errors of reserves as an instrument in equation (5). This identification strategy is inspired by Hamilton (1997), who estimates the slope of the reserve demand curve in 1989–1991 using reserve forecast errors from a time-invariant model as instrument.

The idea is to use variation in reserves that is residual to the Federal Reserve response to dislocations in the federal funds rate and uncorrelated with the transient confounding factors due to the Federal Reserve non-reserve liabilities. Focusing on the main liabilities, the balance-sheet identity for the Federal Reserve implies that

$$\text{Assets}_t^{Fed} = \text{Reserves}_t + \text{TGA}_t + \text{ONRRP}_t + \text{Other}_t,$$

where Assets_t^{Fed} represents the Federal Reserve assets on day t , and Other_t includes non-reserve liabilities other than the TGA and ONRRP (e.g., currency in circulation).²² As a result, reserves change either because the Federal Reserve expands or contracts its balance sheet by trading with or lending to banks, or because the non-reserve liabilities change through transactions with the banking system (see Section 2.3). The total supply of reserves on day t is therefore the sum of the following four components:

$$q_t = q_t^{Fed} + q_t^{TGA} + q_t^{ONRRP} + q_t^{Other},$$

where changes in non-reserve liabilities cause opposite changes in reserves.

To identify the elasticity β_t , we need variation in reserve supply that is exogenous to bank demand at time t . To understand our strategy, it is useful to write the total supply as

$$q_t = H_t + v_t^s, \tag{6}$$

where H_t is the systematic component, which depends on past q_t and p_t , and v_t^s are daily shocks, which are serially uncorrelated (being the residual of the systematic part) but can have stochastic volatility. Both H_t and v_t^s reflect both the actions of the Federal Reserve

²²Other non-reserve liabilities include deposits of foreign central banks, international organizations, Designated Financial Market Utilities (DFMUs), and government-sponsored enterprises (GSEs).

and activity in non-reserve liabilities. Since some fluctuations in non-reserve liabilities are temporarily correlated with bank demand for overnight funding, v_t^s can be correlated with the demand shocks v_t^d in equation (5) up to a finite number of days (see Section 2.3).

Our identification strategy is based on the observation that, at the daily frequency, the systematic supply H_t depends on past reserves and federal funds-IORB spreads, but not on the contemporaneous spread. This is because of the following reasons.

First, the Federal Reserve changes the supply of reserves to implement quantitative easing and tightening (QE and QT) and to counteract dislocations in the federal funds rate. QE and QT represent most of the variation in reserves in our sample; they are slow, persistent processes that depend on past levels of reserves and spreads but do not depend on today's spread. Moreover, in the current monetary policy framework, the Federal Reserve responds to unusual federal funds volatility quickly (September 2019, March 2020), but always with a delay of at least one day; it does not respond on the same day.

Second, variation in the TGA reflects changes in the net proceeds from federal taxes and Treasury debt issuance. These fluctuations mainly depend on government spending and macroeconomic conditions. The proceeds from Treasury issuance also depend on investors' demand for Treasuries, as reflected in the discount and coupon rates determined at auction. These rates may depend on other money-market rates, such as the federal funds-IORB spread, and on the supply of other safe assets, such as reserves.²³ The settlement of Treasury auctions, however, happens a few days after the auction, and coupon payments are issued months after settlement. For this reason, today's variation in the TGA can depend on past reserves and federal funds-IORB spreads, but not on today's spread.²⁴

Third, MMFs are the main investors at the ONRRP, representing 90% of the facility usage since its introduction in 2013. MMF ONRRP investment reflects both their portfolio composition and their size. MMF portfolio composition depends on the rates of the assets that these institutions can hold; since MMFs can neither invest in the federal funds market nor earn the IORB, the share of their portfolio invested at the ONRRP is not a function of the federal funds-IORB spread. MMF size depends on their past net yields relative to those of similar products such as bank deposits because investors respond to past fund performance (La Spada, 2018), but not on today's federal fund-IORB spread.²⁵ MMF size can also depend

²³Some Treasury securities have a floating rate based on money-market conditions; this rate, however, depends on past Treasury bill rates, not on the federal funds-IORB spread.

²⁴Although today's TGA does not depend on today's spread, banks' demand for overnight funding, including federal funds, does depend on some TGA fluctuations as we explain in Section 2.3.

²⁵Moreover, most MMF shares settle either at the end of the day (after 4PM) or on the next day; therefore, MMF size at the time of ONRRP investment (12:45PM) can only reflect yesterday's market rates.

on past reserves through banks' balance-sheet costs: large reserve injections tighten banks' balance-sheet constraints, leading them to push depositors into MMFs (Afonso et al., 2022a).

All other non-reserve liabilities, such as currency, are unlikely to depend on the federal funds-IORB spread or reserves. Moreover, except for currency, they are significantly smaller than reserves, the TGA, or the ONRRP (see Figure IA.1 in Internet Appendix A).

Most of the systematic variation in reserve supply in our sample is due to QE and QT, which are slow-moving persistent processes that can depend on past market conditions in a time-varying fashion. For this reason, we approximate H_t with a linear function of past reserves and spreads with daily time-varying coefficients and m lags, so that total supply is

$$q_t = [a_t + b_{q,1,t}q_{t-1} + b_{p,1,t}p_{t-1} + \dots + b_{q,m,t}q_{t-m} + b_{p,m,t}p_{t-m}] + v_t^s. \quad (7)$$

Substituting this expression in our locally linear demand curve (5) and iterating backward, we can write supply and demand as the following system of moving averages

$$\begin{cases} q_t = \gamma_{a,t} + v_t^s + \sum_{j=1}^{t-1} \gamma_{q,s,j,t} v_{t-j}^s + \sum_{j=1}^{t-1} \gamma_{q,d,j,t} v_{t-j}^d + f_q(q_0, p_0), \\ p_t = \alpha_t + \beta_t \left[\gamma_{a,t} + v_t^s + \sum_{j=1}^{t-1} \gamma_{q,s,j,t} v_{t-j}^s + \sum_{j=1}^{t-1} \gamma_{q,d,j,t} v_{t-j}^d + f_q(q_0, p_0) \right] + v_t^d, \end{cases} \quad (8)$$

where the γ 's are functions of lagged α 's and β 's from equation (5) and current and lagged a 's and b 's from equation (7), and $f_q(q_0, p_0)$ is a function of the initial conditions that is known as of time 0.

Our goal is to estimate β_t , the slope of the reserve demand curve, for every t . Of course, we cannot just regress p_t on q_t because demand shocks are likely to be serially correlated and, on some dates, they are contemporaneously correlated with supply shocks.

The idea is to use the forecast error for reserves $\text{FE}_t^q \equiv q_t - E_{t-1}(q_t) = v_t^s$ as instrument in an IV estimation. It is easy to see that this instrument solves the issue caused by serial correlation in demand shocks. If there are short-term correlations between supply and demand shocks, however, using contemporaneous forecast errors leads to biased estimates.

To see this, suppose that $\text{cov}(v_{t'}^d, v_{t'}^s) = \eta \neq 0$ for some t' , and $\text{cov}(v_t^d, v_k^s) = 0$ for all other t and k . We can think of t' as a Treasury settlement day.²⁶ For notational simplicity, we also

²⁶On settlement days, reserves decline because banks transfer reserves to the TGA to pay for securities, while their demand for overnight borrowing increases to finance those purchases. As a result, the surprise component of TGA change is correlated with banks' demand shock.

suppose that $\text{var}(v_t^s) = \sigma_s^2$ for all t and drop the t subscript from all covariances. If we use contemporaneous forecast errors as instruments for reserves, the IV estimate of β_t will be

$$\begin{aligned}\beta_t^{IV} &= \frac{\text{cov}(p_t, \text{FE}_t^q)}{\text{cov}(q_t, \text{FE}_t^q)} = \frac{\beta_t \sigma_s^2}{\sigma_s^2} = \beta_t \quad \text{for all } t \neq t', \text{ and} \\ \beta_{t'}^{IV} &= \frac{\text{cov}(p_{t'}, \text{FE}_{t'}^q)}{\text{cov}(q_{t'}, \text{FE}_{t'}^q)} = \frac{\beta_{t'} \sigma_s^2 + \eta}{\sigma_s^2} = \beta_{t'} + \frac{\eta}{\sigma_s^2} \neq \beta_{t'}.\end{aligned}$$

The IV estimate would be biased on day t' , with the sign of the bias depending on the specific correlation between the non-reserve liability causing the endogeneity issue and the federal funds-IORB spread on day t' . Note that the bias is negligible if the absolute value of the covariance between supply and demand shocks $|\eta|$ is significantly smaller than the variance of supply shocks σ_s^2 ; i.e., these transient confounding factors are small relative to the typical variation in reserve supply.

One way to solve this issue is to use lagged forecast errors. In the example above, for instance, the IV estimates using lagged forecast errors as instruments would be

$$\begin{aligned}\beta_t^{IV} &= \frac{\text{cov}(p_t, \text{FE}_{t-h}^q)}{\text{cov}(q_t, \text{FE}_{t-h}^q)} = \frac{\beta_t \gamma_{q,s,h,t} \sigma_s^2}{\gamma_{q,s,h,t} \sigma_s^2} = \beta_t \quad \text{for all } t \neq t', t' + h, \\ \beta_{t'}^{IV} &= \frac{\text{cov}(p_{t'}, \text{FE}_{t'-h}^q)}{\text{cov}(q_{t'}, \text{FE}_{t'-h}^q)} = \frac{\beta_{t'} (\gamma_{q,s,h,t'} \sigma_s^2 + \gamma_{q,d,h,t'} \eta)}{\gamma_{q,s,h,t'} \sigma_s^2 + \gamma_{q,d,h,t'} \eta} = \beta_{t'}, \text{ and} \\ \beta_{t'+h}^{IV} &= \frac{\text{cov}(p_{t'+h}, \text{FE}_{t'+h}^q)}{\text{cov}(q_{t'+h}, \text{FE}_{t'+h}^q)} = \frac{\beta_{t'+h} (\gamma_{q,s,h,t'+h} \sigma_s^2 + \gamma_{q,d,h,t'+h} \eta)}{\gamma_{q,s,h,t'+h} \sigma_s^2 + \gamma_{q,d,h,t'+h} \eta} = \beta_{t'+h},\end{aligned}$$

for any $h > 0$. More generally, if $\text{cov}(v_t^s, v_{t+h}^d) = 0$ for all t and $h > \bar{h}$, using reserve forecast errors lagged by $h > \bar{h}$ days gives unbiased estimates of β_t for all t .

Our identification strategy is therefore to estimate

$$\beta_t^{IV} = \frac{\text{cov}(p_t, \text{FE}_{q,t-h})}{\text{cov}(q_t, \text{FE}_{q,t-h})}, \quad (9)$$

where FE_{t-h}^q is the forecast error for reserves from model (7) lagged by h days.

4.3 Empirical implementation

To generate the reserve forecast errors from model (7), we estimate a time-varying VAR with $m = 10$ lags and stochastic volatility that includes an equivalent equation for the federal funds-IORB spread; see equations (A.8) in Appendix B. This model and its estimation are

based on Primiceri (2005) and Del Negro and Primiceri (2015).

The use of a VAR is not necessary because the identification strategy only relies on forecasting the systematic part of reserve supply, which can be done with a scalar time series model (e.g., Hamilton, 1997). Using the VAR, however, is convenient for three reasons. First, it directly provides the joint estimates of the time-varying covariances in the IV estimate (9) because these covariances are equal to the h -day-ahead impulse responses of spreads and reserves to the reserve forecast error from the VAR (see Appendix B.1 for details); that is, we do not need to run a two-stage procedure as in the standard two-stage least-squares (2SLS) estimation.²⁷ Second, since the covariance between the spread and the reserve forecast error is estimated from a VAR with 10 lags and stochastic volatility, the confidence bands around our IV estimates are robust to autocorrelation and heteroscedasticity of the demand shocks in equation (5). Third, adding an equation for the spread improves the predictive accuracy of the model, ensuring that the reserve forecast errors are small and therefore that our locally-linear estimation moves slowly along the demand curve.

We estimate the time-varying VAR at the daily frequency using Bayesian methods; each parameter is modeled as a stochastic process. Consistent with our approximate reserve demand curve (5), the basic assumption behind the estimation of the time-varying VAR is that its parameters evolve more slowly than the daily errors. Following Primiceri (2005), we assume that the parameters follow slow-moving random walks, whose innovations are uncorrelated with the VAR forecast errors at all leads and lags. Given the data, we estimate the joint posterior distribution of all the time-varying parameters on each day. For each day, we then have the joint posterior distribution of the numerator and denominator in the IV estimate (9) and therefore of their ratio.

Inference on the IV estimates (9) depends on the choice of the forecast horizon h . There is a clear trade-off between instrument exogeneity and estimate precision. The longer the horizon is, the more plausible the exogeneity assumption is (i.e., that the instrument is not contaminated by transient correlations between demand and supply shocks). A longer horizon, however, implies larger estimate uncertainty. In our empirical implementation, we use $h = 5$ (i.e., one week) because the confounding factors that cause contemporaneous correlations between non-reserve liabilities such as the TGA and the federal fund-IORB spread are typically transient and only last 2–3 days (Section 2.3). Using 5-day lagged errors should therefore give exogenous instruments, while keeping the estimates sufficiently precise.

²⁷The ratio of impulse response functions as an IV estimator is used by Christiano et al. (1999) to estimate the interest elasticity of the money demand. More recently, Del Negro et al. (2020) and Barnichon and Mesters (2021) use a similar approach for the estimation of the Phillips curve.

Finally, equation (7) models the systematic reserve supply H_t only as a function of past reserves and federal funds-IORB spreads. In reality, fluctuations in the Federal Reserve supply, TGA, and ONRRP can also depend on past repo and T-bill rates, which, in turn, can affect the federal funds-IORB spread as we explain in Section 2.3. To control for this possible endogeneity issue that would lead to an omitted variable bias, our robustness checks include lagged values of these money-market rates in model (7); that is, we add these variables to the time-varying VAR that generates the forecast errors and the posterior distributions of the covariances in the IV estimate (9). This procedure is akin to adding controls to both the first and the second stages of a traditional 2SLS regression. The estimation of these trivariate VARs is similar to that of the baseline bivariate model (see Appendix B.1).

4.4 Model predictive accuracy and instrument relevance

The relevance of our instrument stems from the persistence of reserves in our sample and the forecasting accuracy of our model of reserve supply. Over the last 15 years, the Federal Reserve has expanded and contracted the supply of reserves in response to market conditions, through several QE and QT programs. These trends, which last for months and around which higher-frequency events can occur, are well captured by the autoregressive nature of model (7) and its empirical VAR implementation (A.8). The top panel of Figure 3 shows a scatterplot of the five-day-ahead joint forecasts of the federal funds-IORB spread and reserves from the time-varying VAR, together with the realized data; blue dots are for 2010–2014, gray for 2015–mid-March 2020, red for mid-March 2020–2021, and green for 2022–2024. The forecasts are on top of the realized data across all periods. Since reserves are persistent over time, their time-varying covariance with their forecast errors lagged by five days (i.e., the denominator in the IV estimate (9)) is significant throughout our sample (see panel (b) of Figure 5). This covariance measures the relevance of our instrument, akin to the coefficient on the instrument from the first-stage regression in the traditional 2SLS estimation.

An important advantage of our approach relative to ordinary 2SLS is that our inference is automatically robust to weak instruments. The reason is that the Bayesian posterior distribution of the IV estimate (9) already reflects the uncertainty in both its numerator (i.e., the reduced-form coefficient) and denominator (i.e., the first-stage coefficient). If lagged forecast errors are a weak instrument, the distribution of their covariance with reserves (i.e., the denominator) will be concentrated around zero. While this creates problems in the frequentist approach because asymptotic theory relies on the central limit theorem and Normal approximations, it does not in ours: the posterior distribution of the ratio will

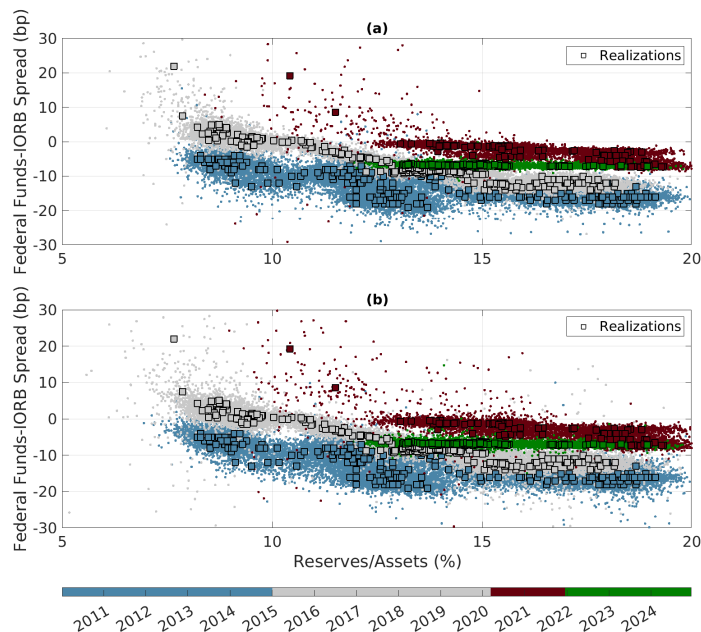


Figure 3: **In-sample (a) and out-of-sample (b) joint forecasts of the federal funds rate and reserves five days ahead.** The forecasts are generated using the bivariate time-varying VAR model (A.8), drawing 100 times every five business days from the joint posterior distribution of reserves and rates for each day. The black squares represent the realized data on the day for which forecasts are generated. Reserves are measured as a ratio to bank assets. The federal funds rate is measured as a spread to the IORB rate.

automatically reflect that the denominator is close to zero with high probability. As a result, we do not need to test for instrument strength or implement robust-inference modifications because our coverage sets already take into account the possible non-normality of the IV estimate caused by instrument weakness.

Importantly, our time-varying VAR displays good predictive accuracy out-of-sample. The bottom panel of Figure 3 replicates the top panel using forecasts generated by estimating the VAR on expanding windows (by 5 days). While the forecast on day t in the top panel uses the in-sample estimates of the model (i.e., all the information in the sample), the forecast in the bottom panel only uses information up to t . The out-of-sample predictions are similar to the in-sample ones and close to the realized data. In Internet Appendix B, we formally evaluate the out-of-sample predictive accuracy of the model, showing that it outperforms univariate and static models. Predictive accuracy is a key advantage of our framework because it enables us to monitor reserve ampleness in real time, as we discuss in Section 5.2.2.

5 Empirical Results

5.1 Demand satiation and the slope of the reserve demand curve

5.1.1 OLS estimation

We now turn to estimating the time-varying slope of the reserve demand curve, i.e., the reserve demand elasticity β_t in equation (5). Before showing the results of our IV estimation, for illustrative purposes, we present the results of a simpler exercise: a rolling-window OLS regression of the federal funds-IORB spread against normalized reserves using in-sample joint forecasts from the time-varying VAR model (A.8) in Appendix B as pseudodata. Every five days, we draw $N = 2,500$ forecasts from the five-day-ahead joint distribution of spreads and reserves and run a pooled regression over the past year (244 days). Figure 4 shows our findings. The slope of the curve changes considerably over time, following the evolution of reserves: it is negative up to mid-2014 as reserves grew from \$1 to \$2.8 trillion, fluctuates around zero between 2014 and 2018 as reserves stayed above \$2 trillion, steadily decreases during 2018-2019 as reserves declined to a minimum of \$1.4 trillion, and moves back towards zero after March 2020 as the Federal Reserve expanded the reserve supply above \$3 trillion.

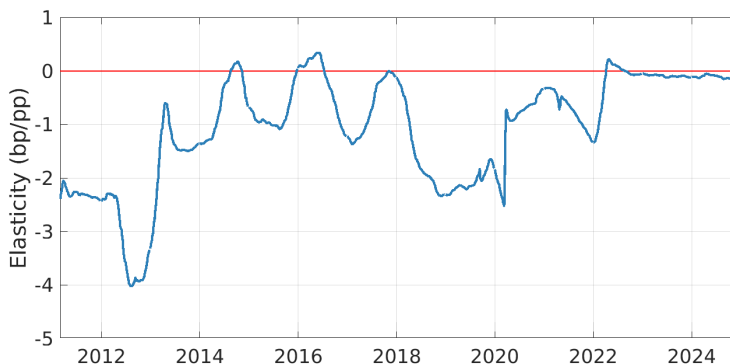


Figure 4: **OLS estimate of the elasticity of the federal funds rate to reserves from model forecast.** The elasticity is estimated by running OLS regressions on rolling windows (244 business days) of in-sample forecasts of the federal funds-IORB spread against in-sample forecasts of reserves normalized by bank assets. The forecasts are generated using the bivariate time-varying model (A.8), drawing 100 times every five business days from the joint posterior distribution of the reserves and rates for each day.

5.1.2 IV estimation

The slopes in Figure 4 cannot be interpreted causally because those OLS estimates do not control for endogeneity. To address identification issues, we use the IV approach described in Section 4; Figure 5 presents the results. Panels (a) and (b) show the time-varying posterior medians of the numerator and denominator of our IV estimate in equation (9), together with their 95% and 68% confidence bands; panel (c) shows the same information for the IV estimate itself.

The numerator in equation (9) is the time-varying covariance between the federal funds-IORB spread and past reserve forecast errors, which can be interpreted as the coefficient from the reduced-form regression of the dependent variable against the instrument in the traditional IV estimation. Panels (a) and (c) of Figure 5 show that, over time, the reduced-form coefficient and IV estimate closely move together, in terms of both sign and statistical significance, supporting the validity of our inference.

The denominator in equation (9) is the time-varying covariance between reserves and their past forecast errors, which can be interpreted as the strength of our instrument. Panel (b) of Figure 5 shows that the 95% confidence bands around this quantity are always above zero, suggesting that our instrument is strong throughout the sample. Moreover, when the instrument is relatively weaker, such as in 2010–2013, this is directly reflected in larger confidence bands around the IV estimate in panel (c). As mentioned in Section 4, the fact that our inference is directly robust to instrument weakness is an important advantage of our methodology relative to traditional IV approaches.

Panel (c) shows that, although different in magnitude, our structural estimates of the rate elasticity to reserve shocks are consistent with the evidence in Figure 4: the time path of the IV estimate is similar to the time-varying slope from the rolling OLS regression on the model forecasts.²⁸ The rate elasticity was significantly negative but steadily increasing from 2010, when reserves ranged between 8% and 10% of bank assets, to 2011, when reserves exceeded 12% of bank assets for the first time in their history. Starting in 2012, with normalized reserves hovering around 12%, the elasticity became insignificantly different from zero and remained so throughout 2013–2017, as normalized reserves ranged from 13% to 19%. In early 2018, a negative slope emerged again, as the Federal Reserve balance-sheet normalization led reserves to drop below 13% of bank assets, reaching a minimum of 8% in September 2019.

²⁸The OLS estimates based on rolling-window regressions are more volatile than the IV estimates. One reason is that the time-varying IV estimates are estimated in a more efficient and general way by allowing the model to weigh past data in an adaptive fashion, rather than fixing the length of the rolling window and giving equal weights within the window.

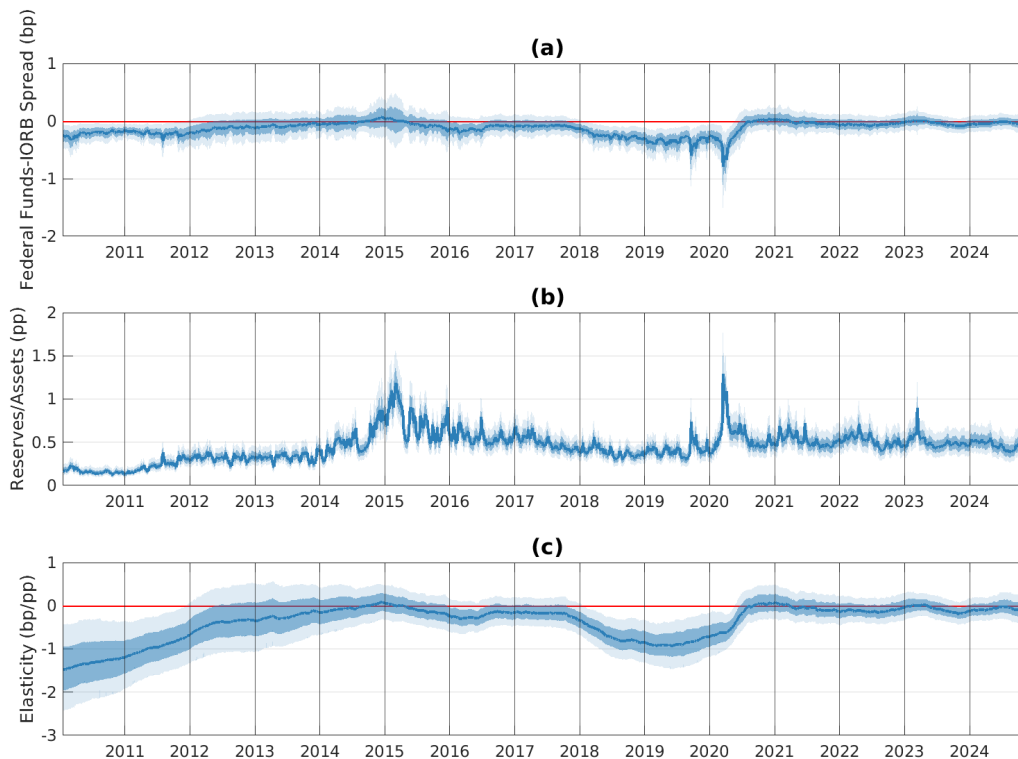


Figure 5: **IV estimate of the elasticity of the federal funds rate to reserves.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample from model (A.8) with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to bank assets, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

The slope returned to be indistinguishable from zero in mid-2020, as the Federal Reserve expanded its balance sheet in response to the Covid crisis and normalized reserves jumped above 16%, staying above 13% through the end of the sample.

Panel (a) of Table 1 reports the quantitative effect of a shock in normalized reserves on the federal funds-IORB spread by year, based on our daily-frequency estimates. For each year, we draw from the joint posterior distribution of the daily IV estimates β_t^{IV} in that year. In 2010, a one-percentage-point drop in the ratio of reserves to bank assets would lead to a median increase in the federal funds-IORB spread of 1.3 basis points. The same drop in normalized reserves would have no effect in 2014; in contrast, it would lead to an increase

of 1 basis point in 2019.

The effects in 2010 and 2019 are also economically important, as they explain a significant share of the in-sample variation in the federal funds-IORB spread. In our sample, the standard deviation of daily changes in the spread is 1 bp; that of daily changes in normalized reserves is 0.2 percentage points (pp). Therefore, our locally linear estimates of the slope of the demand curve imply that, in 2010 and 2019, a daily movement along the curve equal to the standard deviation of reserves' daily changes explains more than 20% of the standard deviation of spreads' daily changes.

Taken together, the results in panel (c) of Figure 5 and in panel (a) of Table 1 are consistent with the nonlinear reserve demand curve with a satiation point predicted by the theory in Section 2.1. Our time-varying estimates of the rate elasticity to reserve shocks suggest that the slope of the reserve demand curve is itself a function of reserves: throughout our sample, it is always significantly negative if the ratio of reserves to commercial banks' assets is below 11%, whereas it is always insignificant if this ratio exceeds 14%.

Importantly, our results are qualitatively similar if we divide reserves by bank deposits or GDP instead of bank assets, confirming that the choice of the normalization factor does not drive our results but simply removes a time trend in reserves; see Internet Appendix C.

Note that the demand satiation implied by our estimates is not driven by the zero lower bound (ZLB). First, the ZLB periods do not coincide with the flat part of the demand curve: rates are at the ZLB in 2010–2011, when the elasticity is significantly negative, while they are above the ZLB in 2016–2017, when the elasticity is zero. Second, the correct dependent variable in our estimation is the federal funds-IORB spread—not the rate itself—because we need to control for changes in the opportunity cost of lending reserves.

Still, for the interpretation of our findings, an important question remains: where does the low-frequency variation in our estimate of the curve's slope come from? The slope of our locally linear approximation can change either because of small exogenous movements along the curve or because of structural horizontal movements of the curve (vertical ones would not change the estimated slope). Since our time-varying estimate closely follows the path of reserves over time, most of the variation in β_t^{IV} seems to come from small exogenous supply shocks captured by our instrument.

Our results, however, also suggest the presence of modest low-frequency horizontal shifts. In fact, the level of reserves at which the demand curve transitions from being flat to being negatively sloped—the satiation level—seems to have changed over time. The transitions between the flat and sloped regions occur when reserves are around 12% of commercial

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
(a) Bi-variate Model															
-1.3	-0.9	-0.4	-0.2	-0.2	0	-0.1	-0.2	-0.2	-0.7	-0.9	-0.2	0	-0.1	-0.1	-0.1
(-2.2,-0.4)	(-1.7,-0.2)	(-1.1,0.4)	(-0.9,0.5)	(-0.5,0.5)	(-0.5,0.4)	(-0.6,0.2)	(-0.5,0.2)	(-1.2,-0.1)	(-1.4,-0.3)	(-1,0.4)	(-1,0.4)	(-0.4,0.3)	(-0.4,0.2)	(-0.4,0.2)	(-0.4,0.3)
(b) Tri-variate Model with Repo Rates															
2	-1.4	-0.9	-0.7	-0.7	-0.2	-0.1	-0.2	-0.2	-0.6	-1	-0.2	-0.1	-0.1	-0.1	-0.1
(-2.9,-1)	(-2.3,-0.5)	(-1.7,0)	(-1.6,0.1)	(-1.1,0.4)	(-0.7,0.4)	(-0.6,0.3)	(-0.7,0.2)	(-0.6,0.2)	(-1.2,-0.1)	(-1.5,-0.4)	(-1.1,0.4)	(-0.5,0.3)	(-0.4,0.3)	(-0.4,0.3)	(-0.4,0.3)
(c) Tri-variate Model with Treasury Yields															
-1.4	-1	-0.5	-0.4	-0.4	-0.1	-0.1	-0.2	-0.1	-0.7	-0.9	-0.2	0	0	0	-0.1
(-2.3,-0.3)	(-1.8,-0.2)	(-1.4,0.4)	(-1.1,0.4)	(-1.1,0.4)	(-0.7,0.5)	(-0.5,0.4)	(-0.7,0.2)	(-0.6,0.3)	(-1.3,-0.1)	(-1.5,-0.3)	(-1,0.5)	(-0.4,0.4)	(-0.4,0.4)	(-0.4,0.4)	(-0.5,0.4)

Table 1: IV estimate of the elasticity of the federal funds rate to reserves by year. The estimate of elasticity is obtained as the posterior median of the ratio between the impulse response of the federal funds rate and the impulse response of reserves to a forecast error in reserves at a five-day horizon; see equation (9). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (A.8); in panel (b), they are estimated from an augmented trivariate version of model (A.8) that also includes daily repo rates; in panel (c), they are estimated in-sample from an augmented trivariate version of model (A.8) that includes daily Treasury yields. All multivariate models include ten lags ($m = 10$ days). The reported elasticities are calculated by year. Reserves are measured as a ratio to bank assets, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

banks' assets in the first half of our sample and around 13% in the second one. These transitions correspond to reserve levels of \$1.5 trillion at the end of 2011 and \$2.2 trillion at the beginning 2018. This modest shift to the right of the demand curve, representing an increase in its satiation point, suggests a modest increase in the demand for reserves. This result is consistent with an increase in bank liquidity regulation and supervision and with Acharya and Rajan (2024), who argue that persistent outward shifts in reserve supply, such as the QE programs of 2010–2014, create outward shifts in reserve demand.²⁹

The results of this section are important for monetary policy implementation because they inform policymakers on the transition between the region of abundant reserves, where the slope of the reserve demand curve is statistically insignificant, and the region of ample reserves, where the slope is significantly negative but only moderately steep. In Section 5.2.2, we also show that, thanks to the predictive accuracy of our forecasting model, our time-varying IV estimates can be used in real time as an early-warning signal of market tightness.

5.1.3 Robustness I: controlling for repo rates and Treasury yields

Model (7) and its empirical time-varying VAR implementation (A.8) assume that the systematic reserve supply only depends on past reserves and federal funds-IORB spreads. Reserves, however, can also depend on other factors, which in turn affect the federal funds-IORB spread, such as repo and T-bill rates (Section 2.3).

To explicitly control for the effect of repo-market conditions, we augment our forecasting model by adding the spread between the daily overnight Treasury repo rate and the IORB rate to the VAR. We then use the reserve forecast errors from this trivariate VAR as instrument for reserves in our IV estimation; as in our baseline specification, we use forecast errors lagged by five days. Adding this possible confounding factor to the VAR is akin to including it as control in both the first- and second-stage regressions of a traditional 2SLS estimation. Figure 6 shows that the results of the trivariate model are consistent with those of the bivariate one: the slope of the reserve demand is negative in 2010–2011 and in 2018–April 2020, whereas it is statistically insignificant during 2012–2017 and after April 2020.³⁰

²⁹In Internet Appendix G, to gauge the possible presence of horizontal and vertical shifts, we run a nonlinear least-squares fit of the time-varying VAR (A.8) forecasts for the federal funds-IORB spread against its forecasts for normalized reserves, allowing for low-frequency horizontal and vertical shifts; this exercise also shows a horizontal outward shift of 2-3 pp from 2010–2014 to 2015–2019.

³⁰The only exception is the 2012Q4–2013Q2 period. Reserves slightly but steadily decline between the second quarter of 2011, just after the second round of large-scale asset purchases, and the fourth quarter of 2012, when the third round starts. As a result, the slope becomes slightly negative in October 2012 but returns to be indistinguishable from zero in June 2013, during the persistent balance-sheet expansion of 2013.

Panel (b) of Table 1 reports the effect of a shock in normalized reserves on the federal funds-IORB spread by year, when controlling for repo rates. Results are quantitatively similar to those from our baseline specification, and even stronger in the earlier part of the sample: a decrease of 1 pp in normalized reserves leads to an increase in the spread of 2 bp in 2010 and 1 bp in 2019, while having no impact during 2012–2017 and 2020–2024.

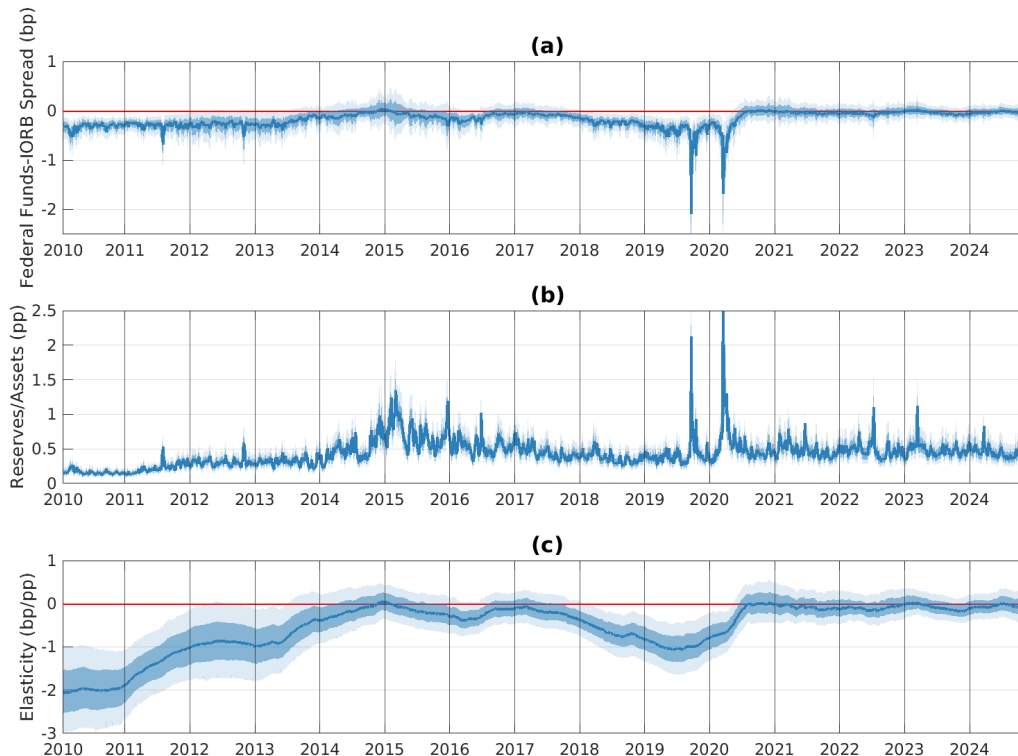


Figure 6: **IV estimate of the elasticity of the federal funds rate to reserves controlling for repo rates.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (A.8) that includes daily repo rates, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to bank assets, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

To explicitly control for the possible confounding effect of Treasury-market conditions, we proceed in a similar fashion. We augment our baseline VAR with the spread between daily market yields on one-year Treasuries and the IORB rate. We then use lagged reserve forecast errors as instrument for reserves in our IV estimation. Figure 7 shows the results:

consistent with the results of the bivariate model in Figure 5 and the trivariate model with repo rates in Figure 6, the demand curve exhibits a negative slope in 2010–2011 and in April 2018–April 2020, while it is flat throughout 2012–2017 and since May 2020.

The estimates of the reserve demand elasticity when controlling for Treasury yields are also quantitatively close to those from the baseline specification. Panel (c) of Table 1 shows that in 2010, a one-percentage-point decrease in normalized reserves leads to a median increase in the federal funds-IORB spread of 1.4 bp. The same drop in the reserves-to-assets ratio has no effect in 2014, whereas it leads to an increase of 1 bp in 2019.

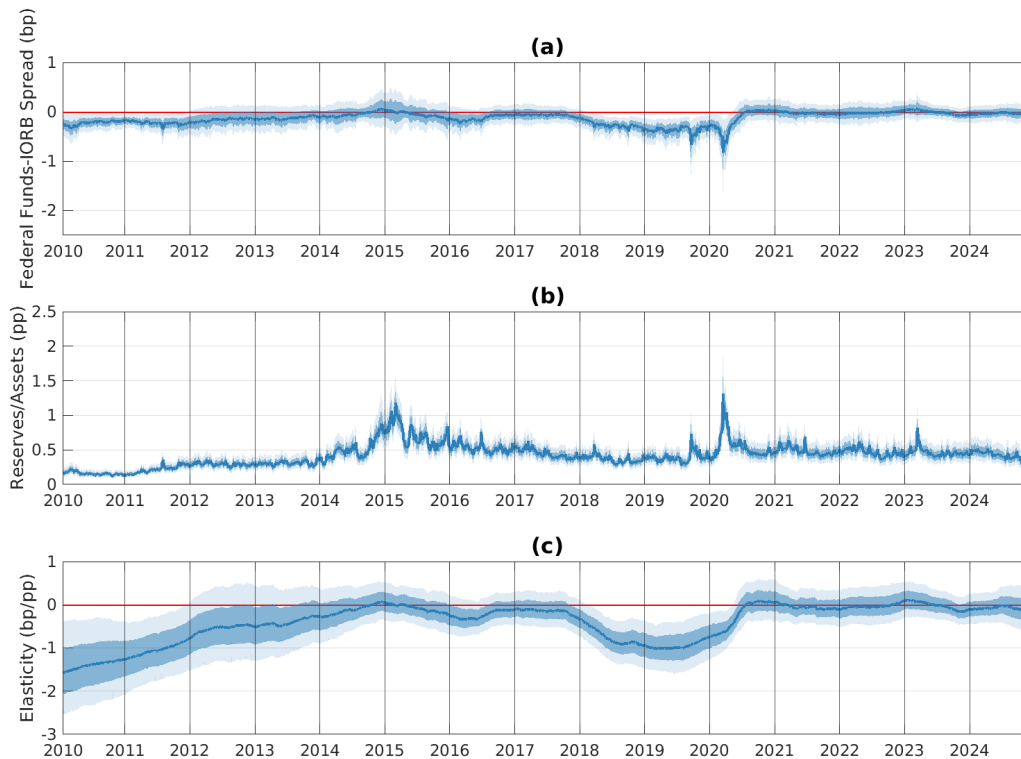


Figure 7: **IV estimate of the elasticity of the federal funds rate to reserves controlling for Treasury yields.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (A.8) that includes daily yields on 1-year U.S. Treasury securities, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to bank assets, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

5.1.4 Robustness II: sign-restricted VAR

The key identifying assumption behind our strategy is that, at the daily frequency, the reserve supply only depends on the past federal funds-IORB spread, not on the current one; as a result, supply shocks can be retrieved through a forecasting model, like our time-varying VAR, without additional structural assumptions. For robustness, we replicate our analysis using a structural time-varying VAR with sign restrictions to extract the supply shocks and covariances that we use in our IV estimation (9).

Sign-restricted VARs impose restrictions on the signs of the contemporaneous effects of the structural shocks on the VAR variables. We implement two variations of this type of structural model. The first one imposes the traditional sign restrictions of supply-demand frameworks: supply shocks have a negative effect on prices and a positive effect on quantities, whereas demand shocks have a positive effect on both. The second variation combines these restrictions with the institutional feature that, in the current monetary policy framework, reserve supply does not contemporaneously respond to demand shocks at the daily frequency; namely, we further assume that the positive contemporaneous effect of demand shocks on quantities is small (i.e., we impose an additional upper bound).

These sign-restricted time-varying VARs are estimated with the same Bayesian method as our baseline model; details on these specifications and their estimation are in Internet Appendix D. Results are in Figures IA.11 and IA.12.

For both robustness checks, the estimates of the reserve demand elasticity follow those of the baseline model: significantly negative in 2010–2011 and 2018–2019, insignificant in the rest of the sample. The estimates based on traditional sign restrictions are more volatile because bivariate VARs of supply and demand identified solely through sign restrictions generate problematic posterior distributions for demand and supply elasticities (Baumeister and Hamilton, 2015; see Internet Appendix D for a detailed discussion). Once we impose that the contemporaneous response of the systematic reserve supply to demand shocks is small, as it the case in the current monetary policy framework, the elasticity becomes smoother and closer in magnitude to the elasticity from our baseline model.

5.2 Validation

5.2.1 Sources of variation in forecast errors

To support the validity of the 5-day-lagged reserve forecast errors from our time-varying VAR as instruments in the demand-elasticity estimation, we show that variation in the errors comes from factors that are uncorrelated with demand for reserves five days later.

Panel (a) of Figure 8 shows a scatterplot of the daily forecast errors for reserves over bank assets from model (A.8) against the contemporaneous daily changes in the TGA over bank assets, together with a fit line. There is a clear negative relationship: a 1 pp daily increase in the ratio between the TGA and bank assets is associated with a contemporaneous forecast error for normalized reserves of -0.63 pp (p -value <0.01 and $R^2 = 0.21$); that is, the model predicts a higher level of reserves than the realized one when the TGA increases, and the size of the error is equal to 63% of the surge in TGA. Panel (b) shows that this negative correlation is even stronger when restricting the sample to the largest 300 forecast errors in absolute value ($\beta = -0.94$, with p -value <0.01 and $R^2 = 0.26$).

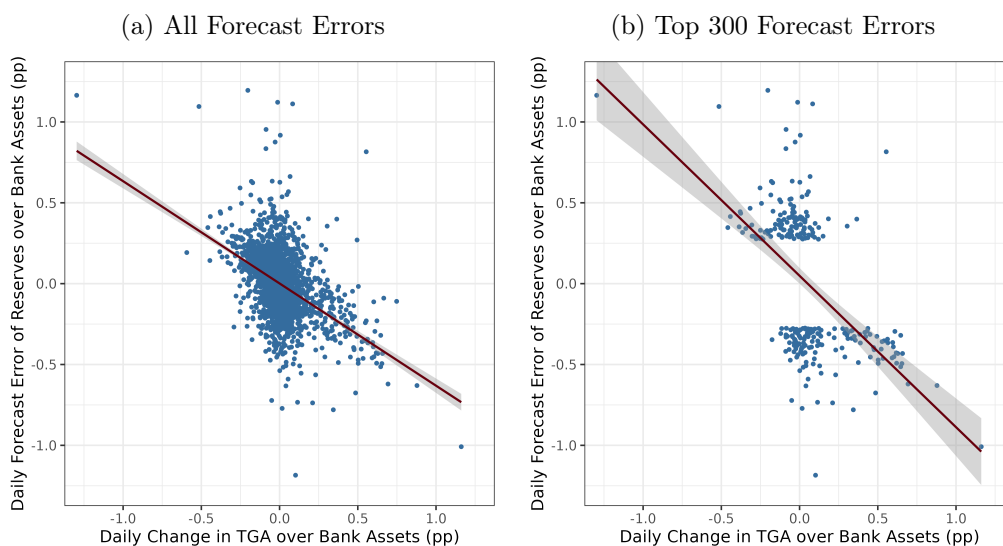


Figure 8: **Daily forecast errors of reserves over bank assets from the time-varying VAR model (A.8) against daily changes in TGA over bank assets, from 2010 to 2024.** Panel (a) includes all forecast errors; panel (b) restricts the sample to the 300 largest errors in absolute value. Each panel includes a linear OLS fit $y = \alpha + \beta x$, with 95% confidence intervals. The results of the OLS fits are: $\beta = -0.63$ (p -value < 0.01) with $R^2 = 0.21$ in panel (a); $\beta = -0.94$ (p -value < 0.01) with $R^2 = 0.26$ in panel (b).

This evidence suggests that variation in our reserve forecast errors comes from unusual

fluctuations in Treasury auctions or tax payments that are not captured by our forecasting model; this result supports the exogeneity of our instrument because, as we explain in the Section 4.2, daily fluctuations in Treasury auctions or tax payments do not depend on contemporaneous changes in the federal funds-IORB spread (i.e., their contribution to the systematic reserve supply is only a function of past spreads) and are uncorrelated with banks' demand shocks after 5 days. Importantly, this evidence is also consistent with the literature that explicitly uses daily TGA fluctuations to estimate the elasticity of the federal funds and other money-market rates to reserve shortages (Hamilton, 1996; Correa et al., 2020).

5.2.2 Out-of-sample performance and real-time monitoring

Our methodology allows us to detect changes in the elasticity of the reserve demand curve—due to either movements along the curve or slow-moving horizontal shifts in the curve—in real time. Figure 9 compares the in-sample (IS) and out-of-sample (OOS) elasticity estimates from our baseline specification. The IS estimate (blue) is the same as the one in Figure 5: for each day, it is obtained using information from the full sample. The OOS estimate (red), in contrast, uses information up to the day at which it is calculated: it is what the econometrician would measure if they re-estimated the model every day, expanding the sample by one day. Comparing IS and OOS estimates shows how our methodology performs in real time.

As Figure 9 shows, IS and OOS elasticities move closely together throughout the sample. As expected, OOS estimates are more volatile and more dispersed at the beginning, reflecting a smaller sample size. As time passes and the sample size increases, the OOS estimates become smoother and their error bands smaller. In 2010 and 2011, the OOS elasticity is negative and significant, with the 95% confidence interval consistently below zero, as with the IS one. During this early period, the OOS estimates are larger in absolute value than the IS ones, likely because the sample size is small and the priors' parameters have been set using data from 2009 (when reserves were scarcer); the difference, however, is never statistically significant at the 95% level. Like the IS elasticity, the OOS elasticity becomes statistically insignificant in 2012 and is practically zero from 2014 throughout 2017.

Albeit with a lag of a few months, the OOS estimates also follow the IS ones in 2018–2019, when the elasticity returned to be negative as reserves steadily declined. In the second half of 2018, the OOS elasticity returns to be significantly negative at the 68% level, and by 2019Q1, it becomes significant at the 95% level and as large as the IS one. That is, if used in real time and taking significance at the 95% level as early-warning signal, our methodology

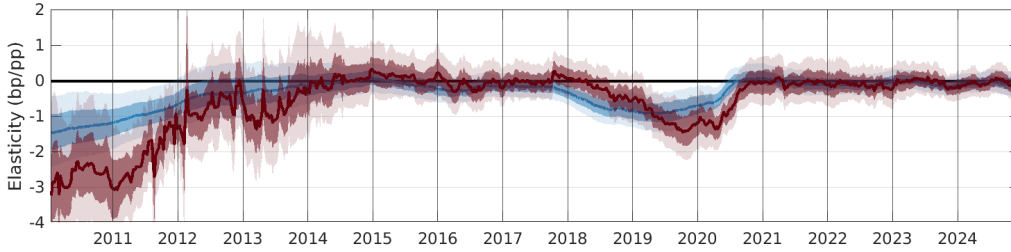


Figure 9: **In-sample (IS) and out-of-sample (OOS) IV estimates of the elasticity of the federal funds rate to reserves.** The estimates are obtained as the ratio between the impulse response of the federal funds-IORB spread and the impulse response of reserves to a forecast error in reserves at a five-day horizon from model (B.1) with ten lags ($m = 10$ days). The blue solid line represents the posterior median IS estimate; the red solid line shows the posterior median OOS estimate. The dark and light shaded areas correspond to the 68% and 95% confidence bands.

would have suggested that we were approaching the negatively sloped region of the reserve demand curve as early as six months ahead of the events of September 2019.

As with the IS elasticity, the OOS elasticity remains significantly negative through June 2020, hovering below the IS one and suggesting that reserves were still relatively scarce in March 2020. The OOS elasticity returns to be insignificant in August 2020, two months after the IS one, and remains practically zero until the end of the sample.

These results show that our methodology can be used in real time to monitor tightness in the federal funds market; a significantly negative elasticity would suggest that even relatively small shocks could lead to significant price dislocations. In particular, we show that our time-varying elasticity estimates would have provided an early-warning signal several months in advance of the events of September 2019.

5.2.3 Other indicators of reserve ampleness

In this section, we compare our estimates of the reserve demand elasticity with four other indicators of reserve ampleness, which are not based on fluctuations in the federal funds-IORB spread. Comparing our estimates with these alternative indicators allows us to check the soundness of our results and validate our methodology.

The first indicator is the daily volume-weighted share of overnight Treasury repos with rates above the IORB. If overnight repo rates are above the IORB (i.e., banks' opportunity cost of lending reserves), banks have an incentive to use their reserve balances to invest in the repo market, thereby keeping repo rates close to the IORB. This reserve-draining

repo intermediation (Correa et al., 2020) is limited by banks' liquidity constraints: if the ampleness of reserves declines, banks will be less willing to drain their reserves, and the share of repo trades above the IORB should increase (d'Avernas et al., 2024). In Internet Appendix E, we also present an alternative measure focusing on quarter-end dynamics.³¹

The second indicator is the daily volume-weighted share of late interbank payments in Fedwire Funds Service.³² Reserves are the settlement asset for interbank payments. As reserves decline and become less ample, banks have an incentive to delay their outgoing payments towards the end of the day to synchronize incoming and outgoing payments and economize on intraday liquidity. As a result, the share of late payments (e.g., after 5 pm) should increase as reserves become less ample (Afonso et al., 2022b; Copeland et al., 2025; Goldstein et al., 2025).

The third indicator is the daily average of the banking system per-minute intraday overdraft. An intraday overdraft occurs when a bank's Federal Reserve account has a temporary negative balance during a business day. Overdrafts are measured at the end of each minute within the business day. To get a sense of the banking system's use of overdrafts, we aggregate overdrafts for all banks at the end of each minute and then average over minutes within the business day.³³ As reserves become less ample and banks more liquidity-constrained, aggregate intraday overdrafts should increase.

Finally, the fourth indicator of reserve ampleness is the daily volume-weighted share of federal funds borrowing by U.S. banks. As opposed to branches of foreign banks, which mainly borrow in the federal funds market for the IORB arbitrage (borrowing below IORB from FHLBs and earning IORB on their reserve balances), U.S. banks typically borrow for liquidity needs. As reserves become less ample, the share of U.S. banks federal funds borrowing should increase.

Figure 10 depicts the 30-business-day moving averages of these additional indicators of reserve ampleness, together with our estimates of the reserve demand elasticity, from 2015 to 2024.³⁴ To have a fair comparison, we plot the OOS elasticity estimates, so that all the

³¹As we explain in Section 2.3, the balance-sheet costs of European banks and their affiliated dealers increase on quarter ends due to the specific implementation of the Basel III leverage ratio in Europe. As a result, intermediation capacity in the repo market declines, and repo rates increase. If reserves are less than ample, the size of these spikes will be larger because U.S. banks will be less willing to use their own reserves to substitute for the decline in European banks' intermediation (Correa et al., 2020; Copeland et al., 2025).

³²Fedwire Funds Service is a real-time gross settlement system that settles transactions individually on an order-by-order basis, without netting. Fedwire Funds Service currently operates twenty-two hours each business day from 9:00 pm Eastern Time (ET) on the preceding calendar day to 7:00 pm ET. From May 2004 to March 7, 2021, Fedwire Funds Service operated twenty-one and a half hours (9:00 pm–6:30 pm ET).

³³The business hours for the calculation of bank overdrafts are the same as those of Fedwire Funds Service.

³⁴We start in 2015 because collection of daily data on the share of federal funds borrowing by U.S. banks

indicators on day t only use information up to that day; since the OOS estimates closely track the in-sample ones, this choice does not significantly affect the comparison. As panels (a)–(d) show, the additional indicators of reserve amplenness negatively comove with our elasticity estimates, supporting the validity of our methodology. Indicators are low in 2015–2017, increase significantly in 2018–2019, and decline again in the second half of 2020. Since then, all the indicators have remained well below their 2018–2019 levels, consistent with our elasticity estimates being statistically insignificant.

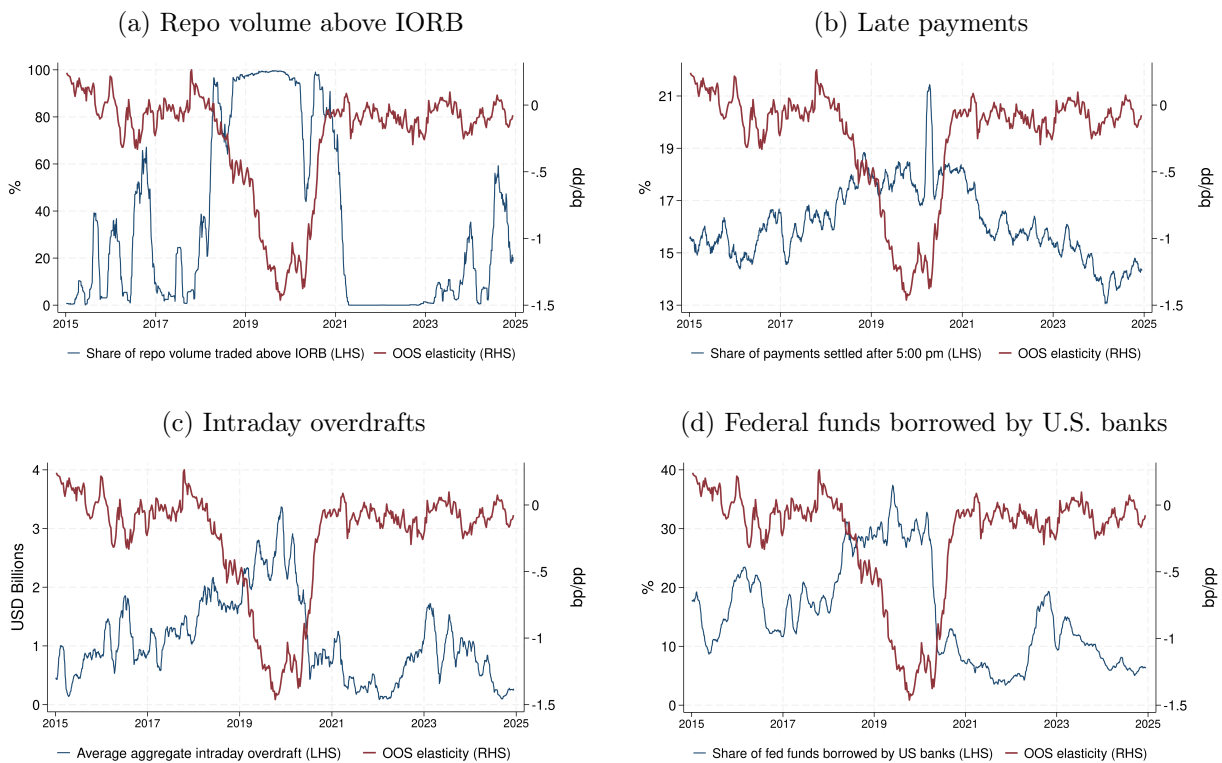


Figure 10: **Additional indicators of reserve amplenness and the OOS elasticity estimates over time.** For each indicator, this figure reports the 30-business-day moving average of its daily value. Panel (a) shows the percentage of overnight Treasury repos with rates above the IORB rate; panel (b) the percentage of Fedwire Funds Service payments settled after 5 pm ET; panel (c) the average aggregate per-minute intraday overdraft in USD billions; panel (d) the percentage of federal funds borrowed by U.S. banks. Percentages are volume-weighted. The red line shows the OOS estimates of the reserve demand elasticity.

This evidence suggests that in addition to externally validating our methodology, these indicators can be used, jointly with our OOS elasticity estimates, as real-time early-warning signals of reserve amplenness during QT cycles. Comparing their values across different peri-

only began in April 2014, and there may have been reporting issues in the earlier months. To be consistent with the rest of the paper, each time series excludes one-day windows around month-ends (see Section 2.3).

ods can inform on changes in the tightness of reserve supply.

Figure 11 shows the values of these indicators and of our OOS elasticity estimates across four different periods between 2015 and 2024, on a spider-web chart. Each variable is normalized such that the inner and outer pentagons correspond to the highest and lowest levels of reserve amplexness between 2015 and 2024, as measured by that variable; e.g., the outer pentagon for the repo indicator corresponds to the share of repos transacted at rates above the IORB in 2019Q3. The periods are chosen to reflect the evolution of the indicators toward a well-known episode of reserve scarcity (September 2019) and a period of reserve abundance (2024Q4).³⁵ All indicators are near the outer edge of the pentagon in 2019 and near the inner edge in 2024. This type of chart can be used as a synoptic tool to monitor reserve amplexness in real time by simultaneously looking at different indicators.

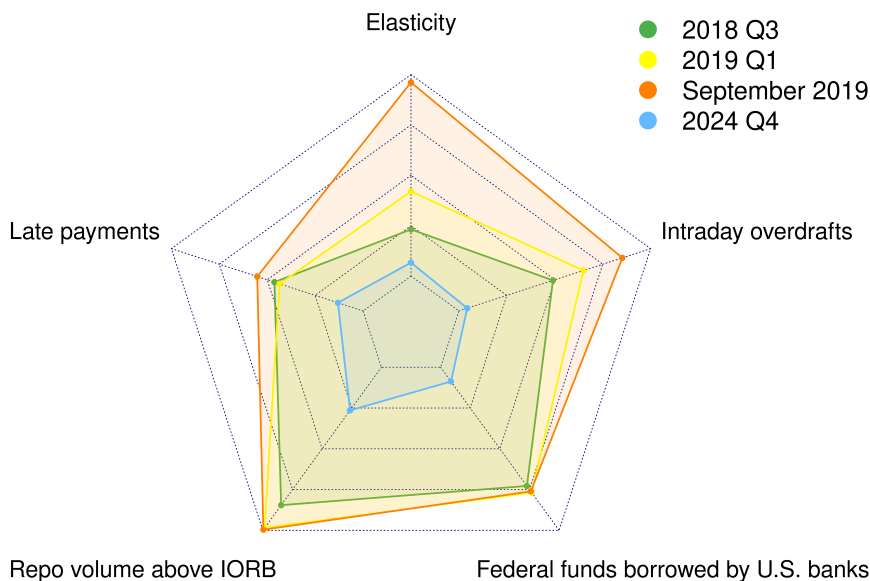


Figure 11: **Indicators of reserve amplexness: spider-web chart.** This chart plots the OOS reserve demand elasticity and four additional indicators of reserve amplexness: (i) the daily share of overnight Treasury repos with rates above the IORB rate; (ii) the daily share of Fedwire Funds Service payments settled after 5 pm ET; (iii) the daily average aggregate per-minute intraday overdraft; (iv) the daily share of federal funds borrowed by U.S. banks. All shares are volume-weighted. Each variable is normalized such that the inner and outer pentagons correspond to the highest and lowest levels of reserve amplexness between 2015 and 2024, as measured by that variable. We show the average values of the normalized variables in: 2018Q3, 2019Q1, September 2019, and 2024Q4.

³⁵2018Q3 and 2019Q1 correspond to the quarters when the OOS elasticity estimate becomes significant at the 68% confidence level and at the 95% confidence level, respectively.

Relative to our elasticity estimates, when used as a monitoring tool, these indicators have the advantage of being easier to calculate. They have, however, two main drawbacks. First, based on economic theory, we can use the reserve-demand elasticity to exactly define the transition between ample and abundant reserves as the level of reserves above which the elasticity is zero (the demand-satiation point). In contrast, there is no similar theory-based threshold for the other indicators of reserve ampleness; they can only be used to make relative statements. This makes the construction of an early warning signal from these indicators more challenging and arbitrary. Second, since these indicators are based on raw data, they are more volatile by construction, which can lead to false positive and false negative signals. The problem is particularly acute for repo rates, as they are affected by several factors other than reserves (e.g., arbitrage opportunities, hedge funds' demand for overnight borrowing). For these reasons, these indicators should be considered in a holistic way when assessing reserve ampleness.

6 Identifying Vertical Shifts

Our estimates of the reserve demand elasticity over time can also help identify the drivers of vertical shifts in the curve. We focus on periods of zero elasticity, when the federal funds market operates in the flat region of the reserve demand curve. During these periods, neither supply shocks nor horizontal demand shifts can affect the federal funds-IORB spread; persistent changes in the spread must be the result of structural vertical shifts.

As we discuss in Section 2.1, two sources of structural vertical shifts in the reserve demand curve are: changes in the spread between the ONRRP and IORB rates and changes in banks' balance-sheet costs. An increase in the ONRRP rate relative to the IORB raises FHLBs' outside option when lending in the federal funds market, pushing rates up; an increase in banks' balance-sheet costs decreases banks' willingness to borrow, pushing rates down. To quantify the effect of changes in the ONRRP-IORB spread, we focus on the adjustments to these rates implemented by the FOMC in 2014–2015; to study the effect of changes in bank balance-sheet costs, we focus on the Supplementary Leverage Ratio (SLR) relief of 2020–2021.

In Internet Appendix F, we repeat our analysis for a few episodes that occurred during periods of negative elasticity (i.e., in the steep region of the curve) and that may have had important effects on the reserve demand curve: the ONRRP-IORB adjustments of 2018–2019 and the 2011 change in the FDIC assessment fee. Results are qualitatively consistent with the ones reported here but smaller in magnitude. Note, however, that since the elasticity is

significantly negative during 2011 and 2018–2019, changes in the federal funds-IORB spread also reflect supply shocks and horizontal demand shifts; the results in Internet Appendix F should therefore be taken cautiously.

6.1 Changes in the ONRRP-IORB spread

6.1.1 Daily horizon analysis

Theory predicts that, when reserves are abundant, the federal funds rate lies below the IORB (i.e., banks’ return from holding reserves) and above the ONRRP rate (i.e., FHLBs’ opportunity cost of lending reserves). An increase in the ONRRP rate relative to the IORB should therefore push the federal funds rate up, closer to the IORB.³⁶ Table 2 shows the list of changes in the ONRRP-IORB spread implemented by the FOMC in our sample.

We focus on the ONRRP-IORB adjustments of 2014–2015 for three reasons. First, during this period, the reserve demand elasticity was zero, ensuring that any change in the federal funds-IORB spread could not be due to supply shocks or horizontal demand shifts. Second, these adjustments were part of an operational readiness exercise, not a response to changing market conditions. Third, the ONRRP volume in 2014–2015 was robust (\$120 billion per day on average), indicating active usage by market participants; in our robustness checks, however, we also include the adjustments of September–December 2013, when ONRRP usage was still low (\$15 billion per day on average).

We run the following time series regression at daily frequency from January 2014 to December 2017:³⁷

$$\begin{aligned} \Delta p_{t+h,t-1} = & \alpha + \beta \Delta \text{ONRRP-IORB}_{t,t-1} + \gamma_1 \Delta \text{Controls}_{t+h,t-1} \\ & + \gamma_2 \text{Calendar Controls}_{t+h} + \varepsilon_{t+h}, \end{aligned} \tag{10}$$

where $\Delta p_{t+h,t-1}$ is the change in the federal funds-IORB spread between day $t - 1$ and day $t + h$, and $\Delta \text{ONRRP-IORB}_{t,t-1}$ is the change in the ONRRP-IORB spread between day $t - 1$

³⁶In general, the pass-through will be smaller than one because the federal funds rate is bounded by the Federal Reserve administered rates. To see this, suppose that reserves are abundant and that the ONRRP rate is increased to be equal to the IORB; since the federal funds rate must lie within these two rates in the region of abundant reserves, it will increase by a smaller amount than the ONRRP rate. The pass-through is also likely to depend on the level of reserves in the system, with the effect being stronger for larger reserve levels: in the region of abundant reserves, the federal funds rate is “anchored” to the IORB and ON RRP rates; in the region of scarce reserves, in contrast, it is also affected by changes in the DW rate.

³⁷We end our sample on December 2017 because the reserve demand elasticity returned to being significantly negative in early 2018.

Date	ONRRP-IORB Spread (bp)		
	$t - 1$	t	Δ
21 Oct. 2013	-24	-23	1
04 Nov. 2013	-23	-22	1
12 Nov. 2013	-22	-21	1
19 Nov. 2013	-21	-20	1
23 Dec. 2013	-20	-22	-2
18 Feb. 2014	-22	-21	1
26 Feb. 2014	-21	-20	1
03 Nov. 2014	-20	-22	-2
17 Nov. 2014	-22	-18	4
01 Dec. 2014	-18	-15	3
15 Dec. 2014	-15	-20	-5
17 Dec. 2015	-20	-25	-5
14 Jun. 2018	-25	-20	5
20 Dec. 2018	-20	-15	5
02 May 2019	-15	-10	5

Table 2: **List of changes to the spread between the ONRRP and IORB rates implemented by the FOMC from September 23, 2013 (when the ONRRP facility was introduced) to December 13, 2024.**

and day t . β measures the change in the federal funds-IORB spread over $h + 1$ days around a change in the ONRRP-IORB spread. We use $h = 0$ and $h = 1$, i.e., we consider one-day and two-day changes in the federal funds-IORB spread.

Since the reserve demand elasticity is zero, neither supply shocks nor horizontal demand shifts affect the federal funds-IORB spread. The spread, however, could still be affected by other factors. To control for high-frequency variation in risk aversion, credit risk, and interest rate risk, we include the $(h + 1)$ -day changes in the VIX, TED spread, and MOVE index. We also include mid-month dummies to control for calendar effects related to the settlement of Treasury auctions and tax payments, day-of-week fixed effects to control for calendar fluctuations in payment activity, and month fixed effects to control for low-frequency time trends.³⁸ Consistent with our estimation of the reserve demand elasticity, we drop one-day windows around month ends to control for the window-dressing practices of European banks (Section 2.3). We use Newey-West standard errors with 5 lags.

Results of regression (10) are in Columns (1) and (2) of Table 3. The pass-through of the

³⁸There are two mid-month dummies. The first captures the change from $t - 1$ to mid-month, i.e., it is equal to one at $t + h = T$, where T is mid-month. The second captures the change from mid-month to $h + 1$ days afterwards, i.e., it is equal to one at $t - 1 = T$.

2014–2015 changes in the ONRRP-IORB spread to the federal funds-IORB spread is 37% on the day of the adjustment and 35% on the following day (p -values < 0.01). In Columns (3) and (4), we replicate the analysis starting the sample in September 2013 to include the adjustments implemented in the early months of the ONRRP facility; results are slightly smaller but similar: 32% at the one-day and 31% at the two-day horizons (p -values < 0.01).

	(1)	(2)	(3)	(4)
	$\Delta p_{t,t-1}$	$\Delta p_{t+1,t-1}$	$\Delta p_{t,t-1}$	$\Delta p_{t+1,t-1}$
$\Delta \text{ONRRP-IORB}_{t,t-1}$	0.37*** (3.70)	0.35*** (4.17)	0.32*** (3.23)	0.31*** (3.53)
R^2	0.13	0.15	0.11	0.13
Sample	1/2014–12/2017	1/2014–12/2017	9/2013–12/2017	9/2013–12/2017
Observations	829	825	883	880

Table 3: **One-day and two-day impact of ONRRP-IORB spread adjustments on the federal funds-IORB spread when reserves are abundant (i.e., flat region of the demand curve).** The table shows the results of regression (10). Data are daily. $\Delta p_{t+h,t-1}$ is the change in the federal funds-IORB spread between $t - 1$ and $t + h$, with $h = 0$ in Columns (1) and (3) and $h = 1$ in Columns (2) and (4). $\Delta \text{ONRRP-IORB}_{t,t-1}$ is the change in the ONRRP-IORB spread between $t - 1$ and t . All columns include the following controls: the changes in the VIX, TED spread, and MOVE index between $t - 1$ and $t + h$ ($h = 0$ in Columns (1) and (3), and $h = 1$ in Columns (2) and (4)); two mid-month dummies (one for $t + h = T$ and one for $t - 1 = T$); day-of-week fixed effects; and month fixed effects. Newey-West standard errors with 5 lags are in parentheses. ***, **, *, represent statistical significance at the 99%, 95%, and 90%, respectively.

6.1.2 Monthly horizon analysis

Since changes in the ONRRP-IORB spread are low-frequency events that permanently change FHLBs’ opportunity cost of lending reserves relative to banks’ benefit from holding them, their effect on the federal funds-IORB spread should be persistent. Moreover, it may take more than a few days for the federal funds rate to react. To gauge the longer-term effects of ONRRP-IORB rate adjustments, we look at their impact at a monthly horizon.

For this analysis, we focus on three episodes: the December 2013 and December 2015 adjustments, because there were no other changes in the ONRRP-IORB spread for one month before and one month after these events; and the February 2014 adjustments. For simplicity, since the two February 2014 changes were close in time (February 18 and 26) and in the same direction (+1 bp in both cases), we treat them as a single event happening on

the day of the first adjustment; results do not change if we use the exact cumulative change in the ONRRP-IORB spread across those days.

For each of the three ONRRP-IORB adjustments, we run the following daily regression on the two-month period surrounding the event (20 business days before and after):

$$p_t = \alpha + \sum_{i=1}^4 \beta_i \mathbf{1}[t \in \text{Post-Change Week}_i] \times \Delta\text{ONRRP-IORB} + \gamma \text{Controls}_t + \varepsilon_t, \quad (11)$$

where p_t is the federal funds-IORB spread, $\text{Post-Change Week}_i$ is a dummy for the i -th week (5 business days) after the adjustment, and $\Delta\text{ONRRP-IORB}$ is the size of the adjustment. β_i measures the average pass-through of the ONRRP-IORB adjustment to the federal funds-IORB spread in the i -th week after the adjustment, using the average spread in the four weeks before the adjustment as control level. The set of Controls include the VIX, TED spread, MOVE index, and day-of-week fixed effects. Since regression (11) is estimated on a two-month window around the event, with the first month serving as control period, we do not include month fixed effects and mid-month dummies.

Results are in Figure 12, with 90%-level confidence intervals based on Newey-West standard errors with five lags. The one-week pass-through of ONRRP-IORB adjustments to the federal funds-IORB spread is between 25% (December 2015; p -value = 0.22) and 52% (February 2014; p -value = 0.05), quantitatively consistent with the estimates of the daily-frequency regressions in Table 3. At the one-month horizon, the pass-through ranges between 64% (February 2014; p -value < 0.01) and 106% (December 2013; p -value < 0.01).

In Internet Appendix F, we replicate our analysis including the changes in the ONRRP-IORB spread of 2018–2019. We show that the pass-through of those adjustments was significantly smaller, consistent with the effect of ONRRP-IORB spread adjustments on the federal funds-IORB spread being weaker when reserves are less ample and rates above the IORB rate.

6.2 Balance-sheet costs: the SLR relief of 2020–2021

Theory predicts that a decrease in bank balance-sheet costs should raise the federal funds rate because banks are willing to borrow at higher rates to expand their balance sheets; as a result, the federal funds-IORB spread should increase (see the model in Appendix A). The SLR relief of 2020–2021 temporarily lowered banks’ balance-sheet costs by excluding reserves and Treasuries from the calculation of the SLR. For depository institutions, those that can

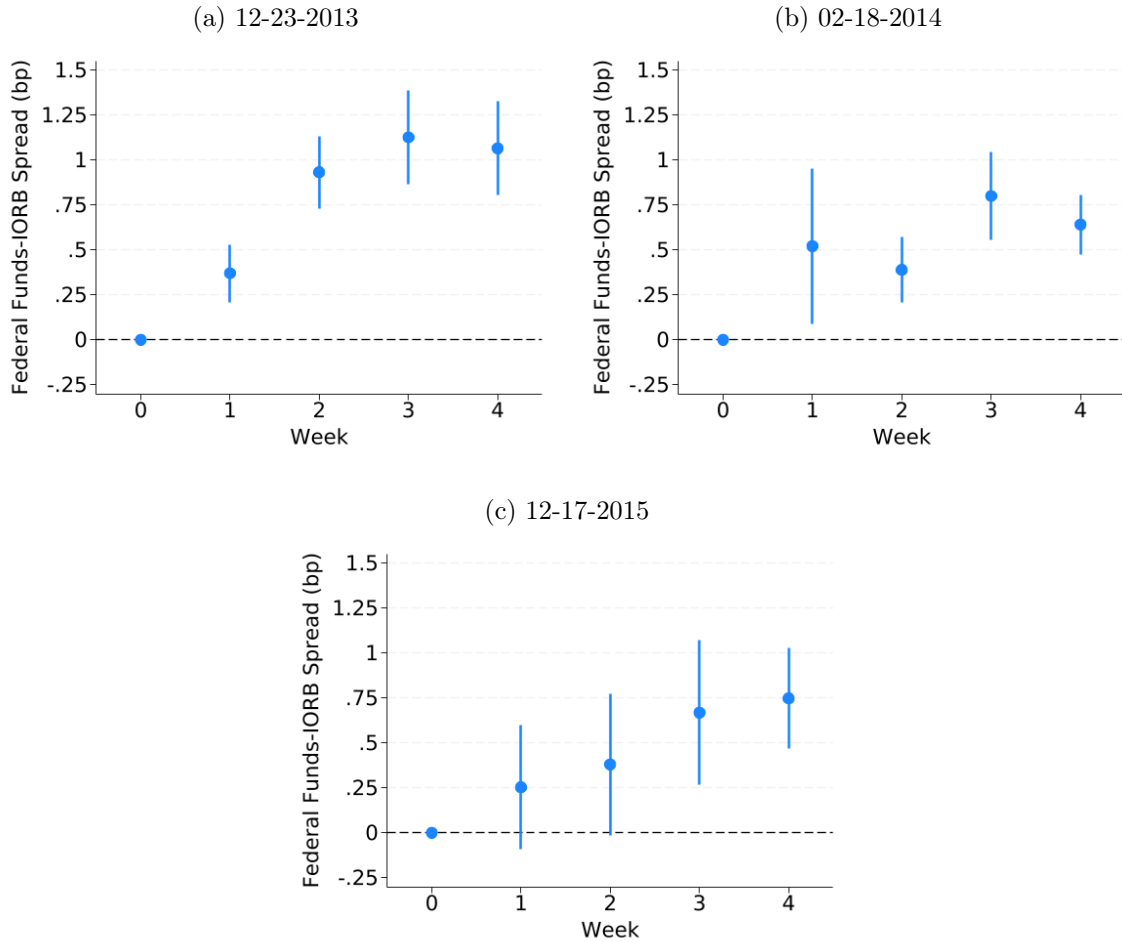


Figure 12: **Monthly horizon impact of ONRRP-IORB spread adjustments on the federal funds-IORB spread when reserves are abundant (i.e., flat region of the demand curve).** Each chart shows the estimated β_i from running regression (11) on a 40-day window around a specific ONRRP-IORB rate adjustment (20 days before and after), together with 90% confidence intervals based on Newey-West standard errors with 5 lags. Panel (a) is for the adjustment of December 23, 2013; panel (b) for the adjustment of February 18, 2014; and panel (c) for the adjustment of December 17, 2015.

hold reserves and trade in the federal funds market, the relief was announced on May 15 and became effective on June 1, 2020. The expiration date of the relief was March 31, 2021, and was announced at the time of its introduction.

Figures 13 shows the federal funds-IORB spread around the SLR relief implementation; consistent with the relief shifting the reserve demand curve upward, the spread increased right after the relief introduction and started to decline ahead of its expiration, which had been known since the temporary regulatory change was announced. Since our estimates of the reserve demand elasticity are insignificant from May 2020 onward, these changes in

the federal funds-IORB spread around the SLR relief cannot be due to supply shocks or horizontal shifts in the demand curve. Moreover, from late July 2020 to late January 2021, the federal funds-IORB spread was stable around -1 bp, suggesting that the spread was not significantly affected by other factors during that period.

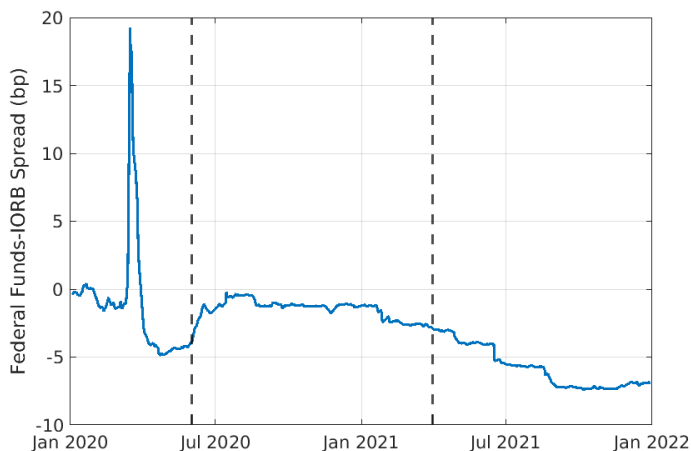


Figure 13: **Federal funds-IORB spread around the implementation of the SLR relief of 2020–2021.** The blue line represents the spread. The two dashed vertical lines represent the introduction (6/1/2020) and end (3/31/2021) of the SLR relief.

To quantify the vertical shift in the reserve demand curve caused by the introduction of the SLR relief, we run the following daily regression over the two months surrounding the event (i.e., four weeks before and after):

$$p_t = \alpha + \sum_{i=0}^N \beta_i * \mathbf{1}[t \in \text{Post SLR-Relief Week}_i] + \gamma * \text{Controls}_t + \varepsilon_t, \quad (12)$$

where $\mathbf{1}[\text{Post SLR-Relief Week}_i]_t$ is a dummy for the i -th week following the SLR relief introduction, and all the variables are defined as in equation (11). As in regression (11), the control period is the four weeks before the event.

Results of regression (12) are in Panel (a) of Figure 14, with 90%-level confidence intervals based on Newey-West standard errors with five lags. Relative to its average level in the previous month, the federal funds-IOR spread increases by 0.6 bp (p -value < 0.01) in the first week after the SLR relief introduction; by the fourth week, the increase is 2.5 bp (p -value < 0.01).³⁹

³⁹This upward shift is consistent with the results in Internet Appendix G, where we run a nonlinear fit of model forecasts allowing for low-frequency shifts across 2010–2014, 2015–2019, and 2020–2024.

We also estimate regression (12) on the eight weeks surrounding the end of the SLR relief, with the week dummies being defined relative to the expiration of the temporary regulatory change. Results are in Panel (b) of Figure 14. The federal funds-IORB spread declines around the end of the relief, but the effect is smaller in magnitude: -0.3 bp within the first week and -0.6 bp at the one-month horizon (p -values < 0.01); the smaller magnitude is consistent with the expiration date being known to market participants long in advance.

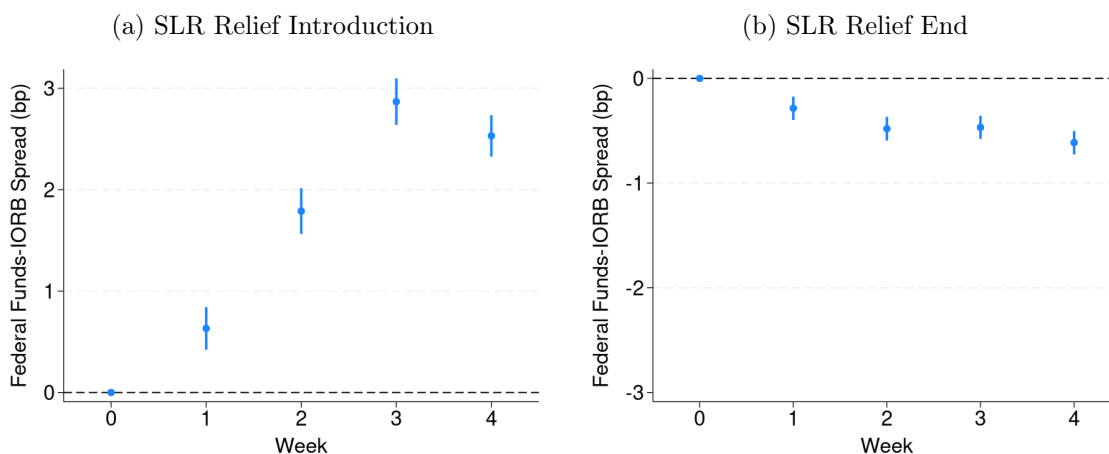


Figure 14: **One-month horizon impact of the 2020–2021 SLR relief on the federal funds-IORB spread.** Panel (a) shows the estimated β_i from running regression (12) on a 40-day window (20 days before and after) around the introduction of the SLR relief (June 1, 2020), together with 90% confidence intervals based on Newey-West standard errors with 5 lags. Panel (b) shows the results from running regression (12) on a 40-day window around the end of the SLR relief (March 31, 2020).

To measure the effect of the SLR relief at longer horizons, we also estimate regression (12) on the four-month periods surrounding the introduction and expiration of the relief (i.e., 40 business days before and after each event). Results are in Figure 15. The effects are stronger at longer horizons, as market adjustments to changes in bank balance-sheet costs may take time. At two months after the introduction of the relief, the federal funds-IORB spread increases by 7 bp (p -value < 0.01) relative to its average level in the two months preceding the relief. As for the regressions on the shorter time period, the effect around the end of the relief is smaller (-2 bp, with p -value < 0.01), consistent with the event being anticipated and market participants acting in advance.

In Internet Appendix F, we replicate our analysis for the 2011 change in the calculation of the FDIC assessment fee, which significantly increased U.S. banks' balance-sheet costs (Banegas and Tase, 2020). Consistent with the predictions of the theory, the implementation of the new fee pushed the federal funds-IORB spread down, representing a downward vertical

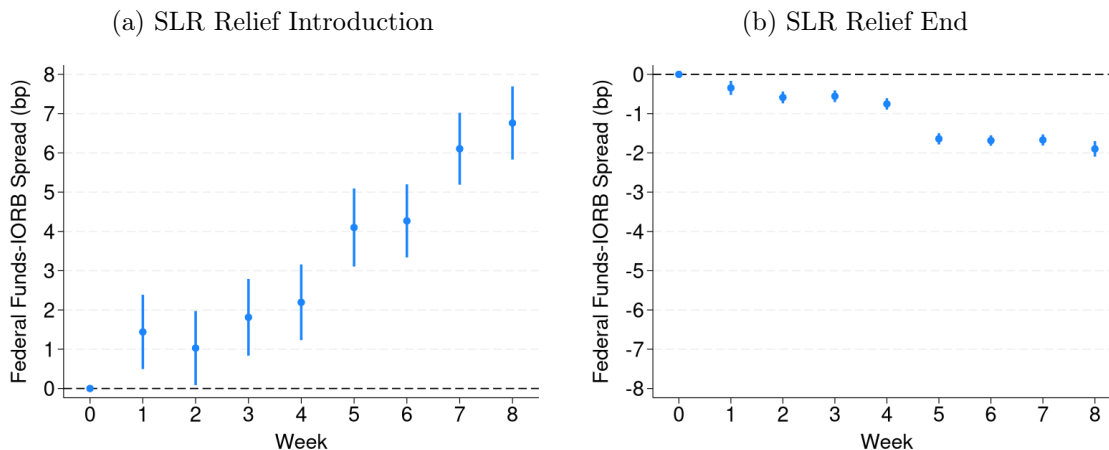


Figure 15: **Two-month horizon impact of the 2020–2021 SLR relief on the federal funds-IORB spread.** Panel (a) shows the estimated β_i from running regression (12) on a 80-day window (40 days before and after) around the introduction of the SLR relief (June 1, 2020), together with 90% confidence intervals based on Newey-West standard errors with 5 lags. Panel (b) shows the results from running regression (12) on a 80-day window around the end of the SLR relief (March 31, 2020).

shift in the reserve demand curve.

7 Conclusion

In this paper, we study the U.S. banking system demand for reserves and identify the level of reserves that satiates it. We provide structural estimates of the different slopes of the reserve demand curve for reserve levels ranging from scarce to abundant, using 15 years of data, from 2010 to 2024.

We find that, as predicted by economic theory, the reserve demand curve is highly nonlinear with a clear satiation level: it is flat when reserves are sufficiently large; and increasingly negatively sloped as reserves decline below the satiation point. Second, we observe horizontal shifts in the demand for reserves. In the earlier part of the sample, we observe a significantly negative slope when reserves are below 12% of bank assets; in the second half, when reserves drop below 13%. These findings suggest that the reserve demand curve, and hence its satiation point, have moved outward over time. Third, we show that the curve has also shifted vertically. Upward vertical shifts are especially relevant in the later part of the sample; this observation has an important implication: the level of the federal funds-IORB spread is not a sufficient statistic for the rate elasticity to reserve shocks. Fourth, using our elasticity

estimates, we identify two drivers of vertical shifts of the reserve demand curve: changes in the spread between the ONRRP and IORB rates and variation in banks' balance-sheet costs.

To produce our time-varying estimates of the reserve demand elasticity, we use an instrumental-variable approach that instruments reserves with past forecast errors from a daily time-varying VAR model of the joint dynamics of reserves and federal funds rates. This methodology addresses the three main issues affecting the estimation of the reserve demand curve: nonlinearity, low-frequency structural changes, and endogeneity. We validate our estimates by comparing them to four alternative indicators of reserve amplex based on conditions in repo markets, timing of payments, banks' borrowing in the federal funds market, and banks' usage of intraday credit. Our findings are robust to normalizing reserves by deposits or GDP and to controlling for changing conditions in repo and Treasury markets. Finally, we also show that our methodology works well out-of-sample and can be used as a tool to monitor reserve amplex in real time.

References

- Acharya, V., Chauhan, R., Rajan, R., and Steffen, S. (2023). Liquidity dependence and the waxing and waning of central bank balance sheets. *NBER Working Paper 31050*.
- Acharya, V. V. and Rajan, R. (2024). Liquidity, liquidity everywhere, not a drop to use – why flooding banks with central bank reserves may not expand liquidity. *Journal of Finance*, 79:2943–2991.
- Afonso, G., Armenter, R., and Lester, B. (2019). A model of the federal funds market: Yesterday, today, and tomorrow. *Review of Economic Dynamics*, 33:177–204.
- Afonso, G., Cipriani, M., Copeland, A., Kovner, A., La Spada, G., and Martin, A. (2021). The market events of mid-September 2019. *Federal Reserve Bank of New York Economic Policy Review*, 27(2).
- Afonso, G., Cipriani, M., and La Spada, G. (2022a). Banks’ balance-sheet costs, monetary policy, and the ON RRP. *Federal Reserve Bank of New York Staff Reports No. 1041*.
- Afonso, G., Duffie, D., Rigon, L., and Shin, H. S. (2022b). How Abundant Are Reserves? Evidence from the Wholesale Payment System. Staff Reports 1040, Federal Reserve Bank of New York.
- Afonso, G., Kim, K., Martin, A., Nosal, E., Potter, S., and Schulhofer-Wohl, S. (2020). Monetary policy implementation with an ample supply of reserves. *Federal Reserve Bank of New York Staff Reports No. 910*.
- Afonso, G., La Spada, G., Mertens, T. M., and Williams, J. C. (2023). The optimal supply of central bank reserves under uncertainty. *Federal Reserve Bank of New York Staff Reports No. 1077*.
- Afonso, G. and Lagos, R. (2015). Trade dynamics in the market for federal funds. *Econometrica*, 83:263–313.
- Anderson, A. G., Du, W., and Schlusche, B. (2020). Arbitrage capital of global banks. *NBER Working Paper 28658*.
- Arias, J. E., Rubio-Ramírez, J. F., Shin, M., and Waggoner, D. F. (2025). Inference based on time-varying SVARs identified with sign restrictions. *Review of Economic Studies*. Forthcoming.

- Armantier, O., Ghysels, E., Sarkar, A., and Shrader, J. (2015). Discount window stigma during the 2007–2008 financial crisis. *Journal of Financial Economics*, 118:317–335.
- Armenter, R. and Lester, B. (2017). Excess reserves and monetary policy implementation. *Review of Economic Dynamics*, 23:212–235.
- Banegas, A. and Tase, M. (2020). Reserve balances, the federal funds market and arbitrage in the new regulatory framework. *Journal of Banking and Finance*.
- Barnichon, R. and Mesters, G. (2021). The Phillips multiplier. *Journal of Monetary Economics*, 117(C):689–705.
- Bassi, C., Behn, M., Grill, M., and Waibel, M. (2024). Window dressing of regulatory metrics: Evidence from repo markets. *Journal of Financial Intermediation*, 58:101086.
- Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999.
- Bech, M. and Klee, E. (2011). The mechanics of a graceful exit: Interest on reserves and segmentation in the federal funds market. *Journal of Monetary Economics*, 58:415–431.
- Benigno, G. and Benigno, P. (2022). Managing monetary policy normalization. *CEPR Discussion Paper No. 17290*.
- Bernanke, B. S. and Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *The American Economic Review*, 82:901–921.
- Bernanke, B. S. and Mihov, I. (1998a). The liquidity effect and long-run neutrality. *Carnegie-Rochester Conference Series on Public Policy*, 49:149–194.
- Bernanke, B. S. and Mihov, I. (1998b). Measuring monetary policy. *The Quarterly Journal of Economics*, 113:869–902.
- Bianchi, J. and Bigio, S. (2022). Banks, liquidity management, and monetary policy. *Econometrica*, 90:391–454.
- Bigio, S. and Sannikov, Y. (2021). A model of credit, money, interest, and prices. *NBER Working Paper 28540*.
- Bindseil, U. (2016). Evaluating Monetary Policy Operational Frameworks. Technical report, Economic Policy Symposium at Jackson Hole, Wyoming.

- Borio, C. (2023). Getting up From the Floor. BIS Working Papers 1100, Bank for International Settlements.
- Christiano, L. J. and Eichenbaum, M. (1992). Liquidity effects and the monetary transmission mechanism. *The American Economic Review*, 82:346–353.
- Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 2, pages 65–148. Elsevier.
- Cipriani, M. and La Spada, G. (2021). Investors’ appetite for money-like assets: The money market fund industry after the 2014 regulatory reform. *Journal of Financial Economics*, 140:250–269.
- Cipriani, M. and La Spada, G. (2022). Implementing monetary policy through non-banks: the ON RRP. *ESCB Legal Conference*.
- Copeland, A., Duffie, D., and Yang, Y. (2025). Reserves were not so ample after all. *Quarterly Journal of Economics*, 140:239–281.
- Correa, R., Du, W., and Liao, G. (2020). U.S. banks and global liquidity. *University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-89*.
- D’Agostino, A., Gambetti, L., and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28:82–101.
- Del Negro, M., Lenza, M., Primiceri, G. E., and Tambalotti, A. (2020). What’s up with the Phillips curve? *BPEA Conference Draft, Spring*.
- Del Negro, M. and Primiceri, G. (2015). Time varying structural vector autoregressions and monetary policy: A corrigendum. *The Review of Economic Studies*, 82:1342–1345.
- Diamond, W., Jiang, Z., and Ma, Y. (2024). The reserve supply channel of unconventional monetary policy. *Journal of Financial Economics*.
- Ding, D., Gonzalez, R. B., Ma, Y., and Zeng, Y. (2025). The effect of instant payments on the banking system: Liquidity transformation and risk-taking. *Working Paper*.
- Duffie, D. (2018). Post-crisis bank regulations and financial market liquidity. <https://www.darrellduffie.com/uploads/1/4/8/0/148007615/duffiebafilecture2018.pdf>.

- d'Avernas, A., Han, B., and Vandeweyer, Q. (2024). Intraday liquidity and money market dislocations. *Working Paper*.
- d'Avernas, A. and Vandeweyer, Q. (2024). Treasury bill shortages and the pricing of short-term assets. *Journal of Finance*.
- Ennis, H. M. and Keister, T. (2008). Understanding monetary policy implementation. *Economic Quarterly*, 94:235–263.
- FOMC (2019). Statement regarding monetary policy implementation. <https://www.federalreserve.gov/newsevents/pressreleases/monetary20191011a.htm>.
- Goldstein, I., Yang, M., and Zeng, Y. (2025). Payments, reserves, and financial fragility. *Working Paper*.
- Greenwood, R., Hanson, S. G., and Stein, J. C. (2016). The Federal Reserve's balance sheet as a financial-stability tool. *2016 Economic Policy Symposium Proceedings*.
- Hamilton, J. D. (1996). The daily market for federal funds. *Journal of Political Economy*, 104(1):26–56.
- Hamilton, J. D. (1997). Measuring the liquidity effect. *The American Economic Review*, 87(1):80–97.
- Hanson, S. G., Scharfstein, D. S., and Sunderam, A. (2015). An evaluation of money market fund reform proposals. *IMF Economic Review*, 63(4):984–1023.
- Kashyap, A. and Stein, J. C. (2012). The optimal conduct of monetary policy with interest on reserves. *American Economic Journal: Macroeconomics*, 4:266–282.
- Keister, T., Martin, A., and McAndrews, J. (2008). Divorcing money from monetary policy. *Federal Reserve Bank of New York Economic Policy Review*, 14:41–56.
- Kim, K., Martin, A., and Nosal, E. (2020). Can the U.S. interbank market be revived? *Journal of Money, Credit and Banking*, 52:1645–1689.
- La Spada, G. (2018). Competition, reach for yield, and money market funds. *Journal of Financial Economics*, 129(1):87–110.
- Lagos, R. and Navarro, G. (2023). Monetary policy operations: Theory, evidence, and tools for quantitative analysis. *NBER Working Paper 31370*.

- Lopez-Salido, D. and Vissing-Jorgensen, A. (2023). Reserve demand, interest rate control, and quantitative tightening. *Federal Reserve Board Working Paper*.
- Martin, A., McAndrews, J., Palida, A., and Skeie, D. (2019). Federal Reserve tools for managing rates and reserves. *Federal Reserve Bank of New York Staff Reports No. 642*.
- Poole, W. (1968). Commercial bank reserve management in a stochastic model: Implications for monetary policy. *The Journal of Finance*, 23(5):769–791.
- Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72:821–852.
- Rossi, B. and Sekhposyan, T. (2019). Alternative tests for correct specification of conditional forecast densities. *Journal of Econometrics*, 208:638–657.
- Rubio-Ramírez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77(2):665–696.
- Schulhofer-Wohl, S. and Clouse, J. (2018). A sequential bargaining model of the fed funds market with excess reserves. *Federal Reserve Bank of Chicago Working Paper No. 2018-08*.
- Smith, A. L. (2019). Do changes in reserve balances still influence the federal funds rate? *Federal Reserve of Kansas City Economic Review*.
- Smith, A. L. and Valcarcel, V. J. (2023). The financial market effects of unwinding the federal reserve’s balance sheet. *Journal of Economic Dynamics and Control*, 146.
- Stein, J. C. (2012). Monetary policy as financial stability regulation. *The Quarterly Journal of Economics*, 127:57–95.
- Swanson, E. T. (2023). The federal funds market, pre- and post-2008. In Gurkaynak, R. S. and Wright, J. H., editors, *Research Handbook of Financial Markets*, Research Handbooks in Money and Finance, chapter 10, pages 220–236. Edward Elgar Publishing.
- Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Appendix

Appendix A Demand for Reserves: Model Extensions

Appendix A.1 Federal Home Loan Banks

In this appendix, we present a variation of the model in Section 2.1 that includes FHLBs. The main difference is that, when lending in the federal funds market, FHLBs have a different outside option than banks: they do not earn i^{IORB} on their balances with the Federal Reserve but can invest at the ONRRP facility to earn $i^{ONRRP} < i^{IORB}$.

To model the interaction between banks and FHLBs, we use the setting of Bech and Klee (2011). Specifically, we make the following assumptions. (i) Consistent with the empirical evidence, FHLBs can only lend in the federal funds market. (ii) When banks borrow in the market, they borrow from another bank with probability $1 - \pi$ or from an FHLB with probability π . (iii) Banks and FHLBs have different outside lending rates: i^{IORB} for banks, and $i^{ONRRP} < i^{IORB}$ for FHLBs. (iv) When banks borrow from FHLBs, the counterparties Nash-bargain over the rate; given the equilibrium rate in the bargaining game, banks then decide how much to borrow, taking into account the uncertainty about the lender type. (v) Finally, we allow banks and FHLBs to have different bargaining powers. See Bech and Klee (2011) for more details.⁴⁰

The bank's optimization problem (1) can then be re-written as

$$\min_{f_i} \left\{ (i^b - i^{IORB}) \int_{\hat{z}_i}^{\bar{z}} (r'_i - \bar{r}_i) g(z) dz + [i^{DW} - (\pi i^h + (1 - \pi) i^b)] \int_{\underline{z}}^{\hat{z}_i} (\bar{r}_i - r'_i) g(z) dz \right\}, \quad (\text{A.1})$$

where i^b is the rate of bank-to-bank lending, i^h is the rate of FHLB-to-bank lending, $i^f = \pi i^h + (1 - \pi) i^b$ is the market rate, and all the other variables are defined as in equation (1). As in the cost function of the baseline model in Section 2.1, the first term represents the opportunity cost of ending the day with a surplus of reserves relative to the target, whereas the second term represents the opportunity cost of ending the day with a deficit of reserves and borrowing from the central bank at the rate i^{DW} instead of borrowing in the market during the day.

⁴⁰For a general framework that incorporates the over-the-counter nature of interbank markets using a search model and that allows banks and FHLBs to negotiate both the rate and size of the loan, see Afonso et al. (2019).

From Bech and Klee (2011), the equilibrium FHLB-to-bank lending rate is

$$i^h = i^b - (1 - \theta) (i^b - i^{ONRRP}), \quad (\text{A.2})$$

where $\theta \in (0, 1)$ is the bargaining power of the FHLBs.⁴¹ The intuition is simple: the FHLB lending rate i^h is smaller than bank lending rate i^b because FHLBs earn a lower rate on their outside option than banks ($i^{ONRRP} < i^{IORNB}$); moreover, i^h increases with FHLBs' bargaining power and with FHLB's outside lending rate.

Plugging equation (A.2) into (A.1), we can rewrite the objective function as

$$\min_{f_i} \left\{ C^{Frictionless} + \pi(1 - \theta) [i^b - i^{ONRRP}] \int_{\underline{z}}^{\hat{z}_i} (\bar{r}_i - r'_i) g(z) dz \right\}, \quad (\text{A.3})$$

where $C^{Frictionless}$ is the same objective function as in the baseline model (see equation (1)), with the bank-to-bank lending rate i^b instead of the overall market rate i^f . Since the market rate i^f in the model with FHLBs is lower than in the baseline model in Section 2.1, the opportunity cost of ending the day with a deficit of reserves is now higher.

Taking the first-order condition of (A.3) with respect to f_i , we obtain

$$i^b = \frac{i^{IORB} + [i^{DW} - i^{IORB} - \pi(1 - \theta) i^{ONRRP}] G(\hat{z}_i)}{1 - \pi(1 - \theta) G(\hat{z}_i)},$$

which can be inverted to obtain the optimal amount of borrowing by bank i :

$$f_i = \bar{r}_i - \tilde{r}_i - G^{-1} \left(\frac{i^b - i^{IORB}}{i^{DW} - i^{IORB} + \pi(1 - \theta) (i^b - i^{ONRRP})} \right),$$

where we plugged in the definition of $\hat{z}_i = \bar{r}_i - \tilde{r}_i - f_i$. Summing across all banks and using the market clearing condition $\sum_i f_i^* = S^h$, where S^h is the total supply of loans from FHLBs, we obtain

$$\bar{r} - \tilde{r} - NG^{-1} \left(\frac{i^b - i^{IORB}}{i^{DW} - i^{IORB} + \pi(1 - \theta) (i^b - i^{ONRRP})} \right) = S^h,$$

where N is the number of banks and the other variables are defined as in Section 2. Solving

⁴¹To derive this relation, Bech and Klee (2011) assume that, after meeting with a FHLB, banks have access to a second round of trading with other banks. As a result, the outside option of banks borrowing from a FHLB is the interbank rate i^b . Since the FHLB-to-bank lending rate is smaller than the bank-to-bank rate, however, bank borrowing in the second round is zero in equilibrium.

for i^b , we obtain

$$i^b = \frac{i^{IORB} + [i^{DW} - i^{IORB} - \pi(1 - \theta)i^{ONRRP}] G\left(\frac{\bar{r} - r}{N}\right)}{1 - \pi(1 - \theta)G\left(\frac{\bar{r} - r}{N}\right)}. \quad (\text{A.4})$$

where $r = \sum_i (\tilde{r}_i + f_i^*) = \tilde{r} + S^h$ is aggregate reserves, including those borrowed from FHLBs, before late-day liquidity shocks are realized. Using (A.2) and (A.4), the aggregate (inverse) demand for reserves is

$$i^f = [1 - \pi(1 - \theta)] \frac{i^{IORB} + [i^{DW} - i^{IORB} - \pi(1 - \theta)i^{ONRRP}] G\left(\frac{\bar{r} - r - S^h}{N}\right)}{1 - \pi(1 - \theta)G\left(\frac{\bar{r} - r - S^h}{N}\right)} + \pi(1 - \theta)i^{ONRRP}. \quad (\text{A.5})$$

It is easy to show that $\lim_{r \rightarrow \infty} i^f = i^{IORB} - \pi(1 - \theta)(i^{IORB} - i^{ONRRP}) \in [i^{ONRRP}, i^{IORB}]$, which is equal to i^{IORB} if there are no FHLBs ($\pi = 0$) or FHLBs have all bargaining power ($\theta = 1$), and to i^{ONRRP} if there are only FHLBs and they have no bargaining power ($\pi = 1$ and $\theta = 0$).⁴² The curve's upper bound is given by $\lim_{r \rightarrow -\infty} i^f = i^{DW}$, as in the frictionless case considered in the paper.

From (A.5), it is also easy to see that i^f increases with the interest rate on FHLBs' outside option i^{ONRRP} and with their bargaining power θ , while it decreases with the share of FHLBs in the market π . In particular, for the curve's lower asymptote, we have the following: $\frac{d}{d i^{ONRRP}} \left(\lim_{r \rightarrow \infty} i^f \right) = \pi(1 - \theta) > 0$, $\frac{d}{d \theta} \left(\lim_{r \rightarrow \infty} i^f \right) = \pi (i^{IORB} - i^{ONRRP}) > 0$, and $\frac{d}{d \pi} \left(\lim_{r \rightarrow \infty} i^f \right) = -(1 - \theta) (i^{IORB} - i^{ONRRP}) < 0$. In Section 6.1, we empirically test that the exogenous changes in the ONRRP rate (relative to the IORB rate) implemented by the Federal Reserve in 2014–2015 vertically shifted the flat part of the reserve demand curve in the abundant reserve region, consistent with these theoretical predictions.

Appendix A.2 Bank Balance-sheet Costs

In this appendix, we present a variation of the model in Section 2.1 that incorporates banks' balance-sheet costs. Banks optimize their end-of-day reserve level as in Section 2.1; the main

⁴²When $\theta = 1$, FHLBs can push the market rate equal to the interest on reserves even if their outside option has a lower rate of return.

difference relative to the baseline model is that banks face a cost proportional to the size of their balance sheets at the end of the day.

As in the baseline model, banks can hold two assets in their balance sheets: reserves and federal funds loans.⁴³ At the end of the day, bank i compares its end-of-day balances r'_i with its target balance \bar{r}_i . If bank i enters the end-of-day period with excess reserves ($r'_i > \bar{r}_i$), the size of its balance sheet is $r'_i - \min\{0, f_i\}$, where $f_i < 0$ ($f_i > 0$) denotes bank i 's lending (borrowing) in the federal funds market. If the bank enters the end-of-day period with a reserve deficit $r'_i < \bar{r}_i$, it borrows at the discount window the amount $\bar{r}_i - r'_i$ necessary to meet its target \bar{r}_i ; in this case, its balance-sheet size will be $\bar{r}_i - \min\{0, f_i\}$. Let $\kappa > 0$ be the balance-sheet cost per unit of assets; for simplicity, we assume κ is the same for all banks.

The bank's optimization problem (1) can then be re-written as

$$\min_{f_i} \left[(i^f - i^{IORB} + \kappa) \int_{\hat{z}_i}^{\bar{z}} (r'_i - \bar{r}_i) g(z) dz + (i^{DW} - i^f) \int_{\underline{z}}^{\hat{z}_i} (\bar{r}_i - r'_i) g(z) dz - \kappa \min\{0, f_i\} \right], \quad (\text{A.6})$$

where $\hat{z}_i \equiv \bar{r}_i - \tilde{r}_i - f_i$, as in equation (1).

Objective function (A.6) is not differentiable due to the minimum function in the last term. To simplify the derivation, we use a smooth approximation of the minimum: $\min\{0, x\} \approx -\frac{1}{\alpha} \ln(1 + e^{-\alpha x})$ for large $\alpha > 0$. An advantage of this smooth minimum is that it is strictly increasing and strictly convex everywhere. Under this approximation, we can simply take the first-order condition for equation (A.6) to derive the optimal federal funds net borrowing:

$$i^f = (i^{IORB} - \kappa) + (i^{DW} - i^{IORB} + \kappa) G(\bar{r}_i - (\tilde{r}_i + f_i^*)) + \frac{\kappa}{1 + e^{\alpha f_i^*}}, \quad (\text{A.7})$$

where f_i^* is the unique minimizer of equation (A.6).

Equation (A.7) represents the inverse demand for reserves of an individual bank in the presence of (smooth) balance-sheet costs proportional to balance-sheet size. It is easy to see that i^f is a smooth decreasing function of f_i^* , converging to $i^{IORB} - \kappa$ from above as f_i^* goes to infinity and converging to $i^{DW} + \kappa$ from below as f_i^* goes to negative infinity.

For a general distribution function G of liquidity shocks, it is not possible to invert (A.7) and use the market clearing condition $\sum_i f_i^* = 0$ to derive the aggregate (inverse) reserve demand curve in closed form. However, since the aggregate curve is the horizontal summation of the individual curves, it retains the properties outlined above, as depicted in Figure 1.

⁴³We can easily generalize this assumption to include loans and other assets.

In particular, the lower asymptote of the aggregate curve is $\lim_{r \rightarrow \infty} i^f = i^{IORB} - \kappa$, where r is aggregate reserves at the beginning of the day. As a result, $\frac{d}{d\kappa} \left(\lim_{r \rightarrow \infty} i^f \right) = -1 < 0$; that is, the vertical location of the flat part of the reserve demand curve in the abundant reserve region strictly decreases with banks' balance-sheet costs. In Section 6.2, we empirically show that the SLR relief of 2020–2021 (which temporarily lowered banks' balance-sheet costs) shifted the lower asymptote of the reserve demand curve upward, consistent with this theoretical prediction.

Appendix B The Time-Varying VAR

Appendix B.1 Model description

To generate daily reserve forecasts, we model the relationship between aggregate reserves and the federal funds rate at a daily frequency using a time-varying vector autoregression (TV-VAR) based on Primiceri (2005) and Del Negro and Primiceri (2015). The model is a multivariate time series model with time-varying coefficients and time-varying covariance matrices for the innovations. The model can be written as follows:

$$\begin{aligned} q_t &= c_{q,t} + b_{q,q,1,t}q_{t-1} + b_{q,p,1,t}p_{t-1} + \dots + b_{q,q,m,t}q_{t-m} + b_{q,p,m,t}s_{t-m} + u_{q,t}, \\ p_t &= c_{p,t} + b_{p,q,1,t}q_{t-1} + b_{p,p,1,t}p_{t-1} + \dots + b_{p,q,m,t}q_{t-m} + b_{p,p,m,t}s_{t-m} + u_{p,t}, \end{aligned} \quad (\text{A.8})$$

where p is the federal funds-IORB spread, q is aggregate reserves divided by banks' total assets, and u_q and u_s are serially uncorrelated, heteroskedastic unobservable errors. These errors are assumed to be jointly normally distributed, with zero mean and a 2×2 covariance matrix Ω_t ; i.e., $(u_{q,t}, u_{p,t})' \sim \mathcal{N}(0, \Omega_t)$ on each day t . The number of lags is $m = 10$.

The vectorized form of model (A.8) is:

$$y_t = c_t + B_{1,t}y_{t-1} + \dots + B_{m,t}y_{t-m} + u_t \quad \text{with } t = 1, \dots, T, \quad (\text{A.9})$$

where y_t is a 2×1 stacked vector of $(q_t, p_t)'$; c_t is an 2×1 vector of stacked constant terms $(c_{q,t}, c_{p,t})'$; $B_{i,t}$, with $i = 1, \dots, m$, are the following 2×2 matrices of time-varying coefficients:

$$B_{i,t} = \begin{bmatrix} b_{q,q,i,t} & b_{q,p,i,t} \\ b_{p,q,i,t} & b_{p,p,i,t} \end{bmatrix}.$$

To model time variation in the covariance matrix of the errors, we reparameterize Ω_t as follows:

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \quad (\text{A.10})$$

where $\Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{bmatrix}$ is a diagonal matrix, and $A_t = \begin{bmatrix} 1 & 0 \\ \alpha_{21,t} & 1 \end{bmatrix}$ is a lower triangular matrix. It follows that

$$\begin{aligned} y_t &= c_t + B_{1,t} y_{t-1} + \dots + B_{m,t} y_{t-m} + A_t^{-1} \Sigma_t \varepsilon_t, \\ \text{Var}(\varepsilon_t) &= I_n, \end{aligned} \quad (\text{A.11})$$

where ε_t is a 2×1 vector of reserve and rate shocks that are uncorrelated with each other at each point in time by construction. The factorization of the covariance matrix in (A.10) is convenient because the first ε error is proportional to the forecast error in reserves ($\sigma_{1,t} \varepsilon_{1,t} = u_{q,t}$). As shown in the next section, this modeling strategy implies that the impulse response functions of q_t and p_t to $\varepsilon_{1,t-h}$ are proportional to the covariances of q_t and p_t with $u_{q,t-h}$.

Stacking all the time-varying coefficients in a vector B_t , we can represent the model in the following companion form:

$$\begin{aligned} y_t &= X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \\ X_t' &= I_n \otimes [1, y_{t-1}', \dots, y_{t-m}'], \end{aligned} \quad (\text{A.12})$$

where \otimes denotes the Kronecker product.

We model the parameters in the following way:

$$B_t = B_{t-1} + \nu_t, \quad (\text{A.13})$$

$$\alpha_t = \alpha_{t-1} + \zeta_t, \quad (\text{A.14})$$

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t, \quad (\text{A.15})$$

where $\alpha_t = \alpha_{21,t}$ is the non-zero off-diagonal term in A_t , and $\sigma_t = (\sigma_{1,t}, \sigma_{2,t})'$ is the 2×1 vector of diagonal terms in Σ_t . B and α are modeled as random walks; σ_t is modeled as a geometric random walk, which belongs to the broader class of stochastic volatility models. All innovations in the model ($\varepsilon_t, \nu_t, \zeta_t, \eta_t$) are assumed to be jointly normally distributed

with covariance matrix

$$V = Var \left(\begin{bmatrix} \epsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t \end{bmatrix} \right) = \begin{bmatrix} I_2 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \quad (\text{A.16})$$

where I_2 is the 2×2 identity matrix, S is the variance of ζ_t , and Q and W are positive-definite matrices.

In our robustness checks, we consider trivariate versions of this TV-VAR model that also include either repo rates or Treasury yields. We augment y_t to become a 3×1 vector of system variables, with the following order: normalized reserves, the repo-IORB spread (or the Treasury-IORB spread), and the federal funds-IORB spread. The vector B_t expands to include the additional auto-regressive parameters and constant. A_t maintains its lower triangular structure, expanding to

$$A_t = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{21,t} & 1 & 0 \\ \alpha_{31,t} & \alpha_{32,t} & 1 \end{bmatrix},$$

so that α_t in (A.14) becomes a 3×1 vector of the stacked parameters of A_t . Σ_t maintains its diagonal structure and expands to include $\sigma_{3,t}$, so that σ_t in (A.15) becomes a 3×1 vector.

The covariance matrix of ϵ_t expands to become I_3 . The covariance matrices of the parameter innovations (Q , S , and W) also expand to account for the additional parameters. We assume S is a block-diagonal matrix, with blocks corresponding to parameters belonging to separate equations:

$$S = \begin{bmatrix} S_{1,1} & 0 & 0 \\ 0 & S_{2,1,1} & S_{2,1,2} \\ 0 & S_{2,2,1} & S_{2,2,2} \end{bmatrix},$$

where $S_{1,1}$ is the variance of the ζ innovation for α_{21} , and the lower block is the covariance of the ζ innovations for $(\alpha_{31}, \alpha_{32})'$.

Appendix B.2 Covariance between errors and observables

In this section, we show how the covariances in equation (9) can be interpreted as the h -day-ahead impulse responses of spreads and reserves to a reserve shock under a Choleski

decomposition with reserves ordered first, such as the factorization in (A.10).

Let n be the number of variables in the system, i.e., two in our case. We rewrite the VAR in the companion form:

$$\underbrace{\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-m+1} \end{pmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{pmatrix} c_t \\ 0_n \\ 0_n \\ \vdots \\ 0_n \end{pmatrix}}_{\mathbf{c}_t} + \underbrace{\begin{pmatrix} B_{1,t} & B_{2,t} & \dots & B_{m-1,t} & B_{m,t} \\ I_n & 0_{n \times n} & \dots & 0_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & I_n & \dots & 0_{n \times n} & 0_{n \times n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{n \times n} & 0_{n \times n} & \dots & I_n & 0_{n \times n} \end{pmatrix}}_{\mathbf{B}_t} \underbrace{\begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \vdots \\ y_{t-m} \end{pmatrix}}_{\mathbf{Y}_{t-1}} + \underbrace{\begin{pmatrix} u_t \\ 0_n \\ 0_n \\ \vdots \\ 0_n \end{pmatrix}}_{\mathbf{u}_t}$$

Define $\mathbf{J} = (I_n \underbrace{0_{n \times n} \dots 0_{n \times n}}_{m-1 \text{ times}})'$; we have $y_t = \mathbf{J}'\mathbf{Y}_t$ and $u_t = \mathbf{J}'\mathbf{u}_t$. Iterating the model backward for h periods, we get:

$$\mathbf{Y}_t = \left(\mathbf{c}_t + \sum_{j=1}^h \prod_{k=1}^j \mathbf{B}_{t-k+1} \mathbf{c}_{t-j} \right) + \left(\mathbf{u}_t + \sum_{j=1}^h \prod_{k=1}^j \mathbf{B}_{t-k+1} \mathbf{u}_{t-j} \right) + \prod_{k=0}^h \mathbf{B}_{t-k} \mathbf{Y}_{t-h-1},$$

and therefore

$$y_t = \left(\mathbf{J}'\mathbf{c}_t + \sum_{j=1}^h \mathbf{J}' \prod_{k=1}^j \mathbf{B}_{t-k+1} \mathbf{c}_{t-j} \right) + \left(\mathbf{J}'\mathbf{u}_t + \sum_{j=1}^h \mathbf{J}' \prod_{k=1}^j \mathbf{B}_{t-k+1} \mathbf{u}_{t-j} \right) + \mathbf{J}' \prod_{k=0}^h \mathbf{B}_{t-k} \mathbf{Y}_{t-h-1}.$$

We can now compute the covariance between the observables and the reserve forecast error, conditional on the model parameters $\Gamma_{1:T} = \{c_t, B_{1,t}, \dots, B_{m,t}, A_t, \Sigma_t; t = 1, \dots, T\}$. As reserves are ordered first in our system, we can write the reserve forecast error as $u_{1,t} = \mathbf{u}'_t \mathbf{j}_1$, where \mathbf{j}_1 is the first column of \mathbf{J} . Also note that $\mathbf{u}_t = \mathbf{J}u_t$, as $\mathbf{J}\mathbf{J}' = I_{nm}$ by construction. Since the forecast errors have zero mean and are serially uncorrelated, we have

$$\begin{aligned} \text{cov}(y_t, u_{1,t-h} | \Gamma_{1:T}) &= \text{E}[y_t u_{1,t-h} | \Gamma_{1:T}] = \mathbf{J}' \prod_{k=1}^h \mathbf{B}_{t-k} \text{E}[\mathbf{u}_{t-h} \mathbf{u}'_{t-h} | \Gamma_{1:T}] \mathbf{j}_1 = \mathbf{J}' \prod_{k=1}^h \mathbf{B}_{t-k} \mathbf{J} \Omega_{t-h} \mathbf{J}' \mathbf{j}_1, \\ &= \mathbf{J}' \prod_{k=1}^h \mathbf{B}_{t-k} \mathbf{J} \Omega_{t-h} \boldsymbol{\iota}_1 \end{aligned}$$

where $\boldsymbol{\iota}_1 = (1 \underbrace{0 \dots 0}_{n-1 \text{ times}})'$. $\Omega_{t-h} \boldsymbol{\iota}_1$ is the first column of the covariance matrix Ω_{t-h} .

The Cholesky factorization (A.10) implies that $\Omega_t \iota_1 = A_t^{-1} \Sigma_t \Sigma_t' (A_t')^{-1} \iota_1 = (A_t^{-1} \iota_1) \sigma_{1t}^2$. As a result,

$$\text{cov}(y_t, u_{1,t-h} | \Gamma_{1:T}) = \left(\mathbf{J}' \prod_{k=1}^h \mathbf{B}_{t-k} \mathbf{J} A_{t-h}^{-1} \iota_1 \right) \sigma_{1t-h}^2. \quad (\text{A.17})$$

For simplicity, to estimate these covariances at day t , we approximate past values of the model parameters with their most recent value; in this way, the matrix product $\prod_{k=1}^h \mathbf{B}_{t-k}$ simply becomes the matrix power \mathbf{B}_t^h . This approximation is valid because, given our priors, the model parameters evolve significantly more slowly than the daily errors (see Appendix B.3), and because we choose a relatively short time horizon ($h = 5$) for the forecast errors used in our IV estimation.

Up to the scaling factor $\sigma_{1,t}^2$, our estimates of the covariances in (A.17) are therefore equal to the h -day-ahead impulse responses of the system variables to the standardized reserve shocks calculated using factorization (A.10); in fact, in the traditional VAR literature, the i -th variable's impulse response to ε_1 after h days, $\frac{\partial y_{i,t+h}}{\partial \varepsilon_{1,t}}$, is estimated with the i -th element of the vector $\mathbf{J}' \mathbf{B}_t^h \mathbf{J} A_t^{-1} \iota_1$. Note also that the scaling factor $\sigma_{1,t}$ is the same for all variables in the system; as a result, our IV estimate (9), obtained as ratio of the covariances in (A.17), is exactly equal to the ratio of the h -day-ahead impulse responses of spreads and reserves to reserve shocks.

Appendix B.3 Priors

We use Bayesian methods to estimate model (A.8). As outlined in Primiceri (2005), we use the following prior densities for the initial states of the time-varying parameters:

$$P(B_0) = N(\hat{B}, 4 \cdot \hat{\Psi}_B),$$

$$P(\alpha_0) = N(\hat{\alpha}, 4 \cdot \hat{\Psi}_\alpha),$$

$$P(\log \sigma_0) = N(\log \hat{\sigma}, I_n),$$

where $N(\mu, \sigma^2)$ denotes a normal density function with mean μ and variance σ^2 , and \hat{B} , $\hat{\alpha}$, $\log \hat{\sigma}$, $\hat{\Psi}_B$, and $\hat{\Psi}_\alpha$ are set using a time-invariant VAR with the same ordering as in (A.10) estimated by OLS on the pre-sample from 01/05/2009 to 01/19/2010, covering $T_0 = 226$ daily observations. The prior means, \hat{B} and $\hat{\alpha}$, are set to the OLS point estimates. The prior variances, $\hat{\Psi}_B$ and $\hat{\Psi}_\alpha$, are set equal to the sampling variances of the OLS point estimates. The prior means of the initial states of the log-volatilities are set to the logarithm of the

standard errors of the OLS residuals.

Following Primiceri (2005), we set the prior densities for Q , S , and W as:

$$P(Q) = IW(\lambda_1^2 \cdot T_0 \cdot \hat{\Psi}_B, T_0),$$

$$P(S) = IW(\lambda_2^2 \cdot 2 \cdot \hat{\Psi}_\alpha, 2),$$

$$P(W) = IW(\lambda_3^2 \cdot 3 \cdot I_3, 3),$$

where $IW(A, df)$ is the inverse-Wishart density function with scale matrix A and degrees of freedom df . Smaller values of λ_i imply less time variation in the dynamic parameters of the model; we set $\lambda_1 = 0.04$, $\lambda_2 = 0.1$, and $\lambda_3 = 0.01$. These tight priors, especially that on Q , ensure that the model parameters move more slowly than the daily errors and liquidity shocks affecting banks' demand for reserves.

The posterior distribution of the parameters and the forecasts are obtained by Montecarlo simulations, as described in Primiceri (2005), D'Agostino et al. (2013), and Del Negro and Primiceri (2015).

Internet Appendix

Internet Appendix A Non-Reserve Liabilities

The variability of the Federal Reserve’s non-reserve liabilities, especially the TGA and ONRRP, has increased significantly over time (Afonso et al., 2020). Figure IA.1 shows that, excluding currency in circulation, non-reserve liabilities amounted to 32% of reserves in January 2010 and to 40% in December 2024, with a peak of 120% in September 2022. The TGA increased from \$90 billion at the end of 2010 to \$711 billion at the end of 2024, with a peak of \$1.8 trillion in July 2020. Since its inception on September 2013, the ONRRP has fluctuated between less than a billion and \$2.4 trillion, reached in September 2022.⁴⁴

Internet Appendix B Time-varying VAR: Out-of-sample Validation

As discussed in Section 4.4, our model does not suffer from the curse of dimensionality in spite of its flexibility and generality. Indeed, the out-of-sample (OOS) predictions are reasonable and compare well with in-sample (i.e. ex post) predictions. This suggests that the model is able to capture in a parsimonious way the salient features of a time-varying market. In this section, we provide additional evidence in support of this claim.

To evaluate the forecasting performance of the bivariate TV-VAR model (A.8), we conduct a series of OOS forecasting exercises and evaluate the model’s predictive accuracy according to various metrics. To construct OOS predictive forecast densities, the model is recursively estimated, and the forecasts are generated, every 5 business days from January 20, 2010 to December 13, 2024, using an expanding window of observations.

For comparison, we also generate OOS forecasts using two time-invariant models: a standard bivariate VAR and a vector of two independent AR processes (one for each series). In both models, the innovations in q_t and p_t are assumed to have zero mean, to be serially uncorrelated, and to be normally distributed (jointly in the VAR and independently in the AR processes). Both models are estimated via OLS on daily data, using a 260-day rolling window to allow their parameters to adapt to a changing environment. As in the TV-VAR, both the VAR and the AR models include ten lags and are estimated every 5 business days.

⁴⁴Afonso et al. (2022a) discuss the drivers of the recent surge in ONRRP investment.

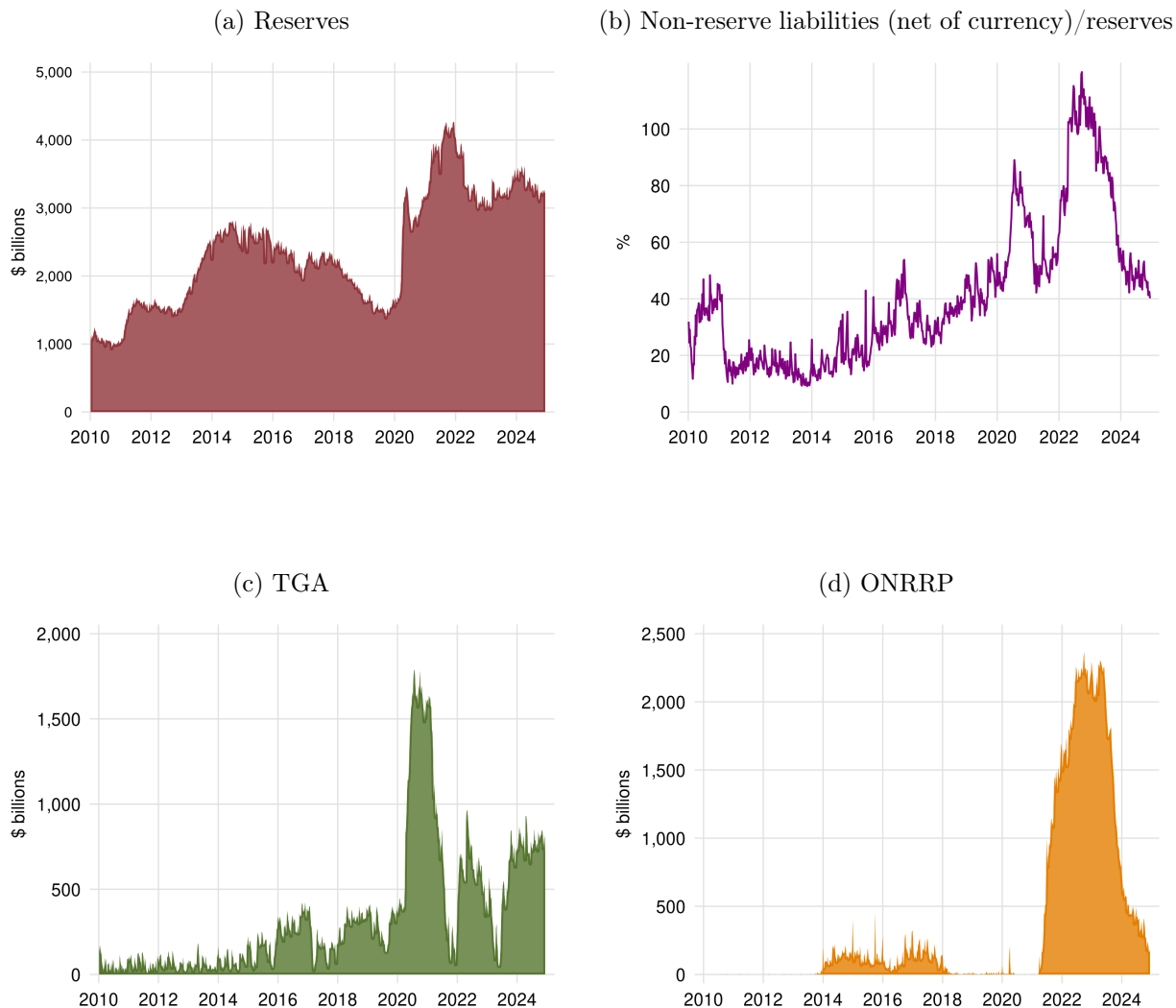


Figure IA.1: **Federal Reserve Liabilities from January 2010 to December 2024.** Reserves (panel (a)), ratio of non-reserve liabilities (excluding currency in circulation) to reserves (panel (b)), Treasury General Account (TGA) (panel (c)), and Overnight Reverse Repurchase Agreement (ONRRP) balances (panel (d)).

Internet Appendix B.1 Point forecasts

We first evaluate the median forecasts for normalized reserves (q_t) and the federal funds-IORB spread (p_t) from the three forecasting models. For each variable we calculate the root-mean-square forecast errors (RMSE) at forecasting horizons of 5, 10, and 20 business days. We also report the determinant of the variance-covariance matrix of the forecast error, as a measure of joint predictive accuracy.

Table IA.1 presents these marginal and joint forecasting performance of each model over

different sample periods. On the full sample, both the marginal and the joint RMSE of the TV-VAR are smaller than those of the VAR and AR models at any forecast horizon. The only exception is the VAR’s RMSE for the spread at $h = 10$, which is equal to the corresponding TV-VAR’s RMSE.

The TV-VAR also displays a higher forecasting accuracy when the RMSE are calculated on non-overlapping two-year windows (e.g., 2010-2011, 2012-2013); only for the 2014-2015 and 2022-2024 periods, when reserves were so abundant that rate fluctuations were unrelated to reserve shocks, are results more mixed, with no model dominating the others.

Internet Appendix B.2 Density forecasts

We then evaluate the entire predictive forecast density. Given the draws from each predictive forecast density, we fit a normal distribution to the marginal draws and a multivariate normal distribution to the joint draws. For each fitted predictive forecast density, we generate a score by evaluating the density function at the realized data and then take its logarithm. This score measures the likelihood of the realized data under the predictive density implied by the forecasting model. Lower scores correspond to lower likelihoods, suggesting lower model accuracy.

Table IA.2 presents the average marginal and joint log scores of each model over the whole sample period and over non-overlapping two-year sub-periods. Over the full sample, the TV-VAR displays significantly higher log-scores both for the marginal density of reserves and for the joint density at all horizons. The log-score for the density of the federal funds-IORB spread is slightly lower than those from the VAR and AR models at horizons of 5, 10, and 20 business days, though it is only materially lower for the latter. When considering the performance by sub-period, all log-scores from the TV-VAR tend to be higher than those from the time-invariant models in the early and late part of the sample (i.e., 2010-2013 and 2018-2021). It is only in the interim (2014-2017) and last (2022-2024) parts, when reserves were so abundant that rate fluctuations were unrelated to reserve shocks, that the TV-VAR’s performance tends to be slightly worse. Unsurprisingly, in that period, univariate models tend to perform slightly better.

Lastly, we assess the calibration of the predictive forecast densities by using probability integral transforms (PITs). The PITs are the values of the predictive marginal cumulative distributions evaluated at the realized data. For each variable, we estimate the PIT by computing the fraction of draws from the forecast density that are less than the realized value. If the predictive density is well calibrated, the PIT should be distributed uniformly

Sample	Model	Reserves/Assets (%)			Federal Funds-IORB Spread (bp)			Joint		
		h = 5	h = 10	h = 20	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
2010 - 2024	TV-VAR	0.412	0.602	0.860	1.50	1.83	2.07	0.607	1.07	1.74
	VAR	0.437	0.645	0.996	1.78	1.82	2.38	0.765	1.14	2.18
	AR	0.424	0.632	1.01	2.38	3.46	5.53	-	-	-
2010 - 2011	TV-VAR	0.342	0.454	0.677	1.47	1.75	2.24	0.442	0.731	1.43
	VAR	0.376	0.526	0.806	1.66	2.09	2.93	0.520	0.899	1.81
	AR	0.358	0.501	0.747	1.58	1.99	2.77	-	-	-
2012 - 2013	TV-VAR	0.338	0.438	0.534	1.27	1.73	2.44	0.416	0.719	1.21
	VAR	0.345	0.473	0.653	1.38	1.88	2.70	0.454	0.799	1.49
	AR	0.346	0.455	0.611	1.37	1.90	2.69	-	-	-
2014 - 2015	TV-VAR	0.584	0.833	1.05	0.779	1.04	1.27	0.454	0.864	1.33
	VAR	0.617	0.774	0.815	0.820	1.04	1.28	0.506	0.807	1.04
	AR	0.588	0.768	0.849	0.778	0.986	1.29	-	-	-
2016 - 2017	TV-VAR	0.364	0.537	0.751	0.424	0.648	0.970	0.152	0.341	0.721
	VAR	0.386	0.611	0.835	0.473	0.714	0.983	0.174	0.416	0.776
	AR	0.374	0.570	0.774	0.462	0.709	1.04	-	-	-
2018 - 2019	TV-VAR	0.262	0.370	0.476	2.44	2.77	2.97	0.556	0.818	1.14
	VAR	0.311	0.455	0.547	3.38	2.70	3.63	0.938	1.04	1.77
	AR	0.270	0.392	0.522	5.43	8.29	14.0	-	-	-
2020 - 2021	TV-VAR	0.484	0.807	1.38	2.49	3.10	3.09	1.20	2.45	4.14
	VAR	0.508	0.922	1.88	2.58	2.84	3.32	1.31	2.58	5.28
	AR	0.520	0.947	1.99	2.67	3.26	3.36	-	-	-
2022 - 2023	TV-VAR	0.450	0.655	0.933	0.229	0.314	0.375	0.102	0.201	0.340
	VAR	0.469	0.690	0.986	0.201	0.260	0.326	0.093	0.179	0.320
	AR	0.463	0.685	0.975	0.199	0.258	0.295	-	-	-
2024	TV-VAR	0.323	0.429	0.469	0.171	0.243	0.280	0.054	0.099	0.120
	VAR	0.331	0.412	0.512	0.160	0.210	0.244	0.053	0.085	0.125
	AR	0.318	0.406	0.497	0.133	0.166	0.182	-	-	-

Table IA.1: **Out-of-sample (OOS) root-mean-square forecast error (RMSE)**. OOS RMSE for normalized reserves and the federal funds-IORB spread from the TV-VAR (A.8) and the VAR and AR models described in Internet Appendix B.

Sample	Model	Reserves/Assets (%)			Federal Funds-IORB Spread (bp)			Joint		
		h = 5	h = 10	h = 20	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
2010 - 2024	TV-VAR	-0.523	-0.874	-1.34	-1.72	-2.06	-3.04	-2.27	-2.94	-4.35
	VAR	-0.657	-1.13	-1.85	-1.70	-1.82	-2.16	-2.54	-3.16	-4.30
	AR	-0.587	-1.05	-1.73	-1.69	-1.85	-2.10	-	-	-
2010 - 2011	TV-VAR	-0.362	-0.653	-1.13	-1.82	-2.02	-2.29	-2.06	-2.63	-3.35
	VAR	-0.570	-0.952	-1.76	-2.02	-2.29	-2.76	-2.40	-3.08	-4.44
	AR	-0.491	-0.857	-1.46	-1.94	-2.19	-2.60	-	-	-
2012 - 2013	TV-VAR	-0.360	-0.656	-0.970	-1.86	-2.20	-2.68	-2.22	-2.89	-3.62
	VAR	-0.424	-0.823	-1.31	-1.75	-2.13	-2.70	-2.13	-2.88	-3.84
	AR	-0.406	-0.735	-1.09	-1.74	-2.14	-2.67	-	-	-
2014 - 2015	TV-VAR	-1.11	-1.44	-1.62	-1.59	-2.46	-4.26	-2.72	-3.93	-5.82
	VAR	-1.25	-1.41	-1.42	-1.30	-1.59	-1.90	-2.64	-3.16	-3.42
	AR	-1.11	-1.39	-1.49	-1.21	-1.47	-1.86	-	-	-
2016 - 2017	TV-VAR	-0.427	-0.836	-1.52	-1.07	-2.24	-4.31	-1.50	-3.07	-5.79
	VAR	-0.465	-0.948	-1.33	-0.550	-0.901	-1.31	-1.04	-1.93	-2.82
	AR	-0.426	-0.835	-1.16	-0.547	-0.890	-1.31	-	-	-
2018 - 2019	TV-VAR	-0.161	-0.532	-0.957	-4.66	-3.27	-3.29	-5.09	-3.76	-4.22
	VAR	-0.342	-0.763	-0.953	-5.56	-4.55	-4.76	-6.91	-6.14	-6.65
	AR	-0.129	-0.515	-0.858	-5.65	-4.90	-4.46	-	-	-
2020 - 2021	TV-VAR	-0.598	-1.04	-2.00	-1.55	-2.03	-3.17	-2.21	-3.12	-5.19
	VAR	-0.998	-2.05	-4.96	-1.94	-2.17	-2.54	-3.32	-4.88	-8.53
	AR	-1.00	-2.07	-4.79	-1.99	-2.36	-2.66	-	-	-
2022 - 2023	TV-VAR	-0.712	-1.03	-1.35	-0.324	-0.928	-1.96	-1.02	-1.95	-3.31
	VAR	-0.706	-1.19	-1.73	0.124	-0.116	-0.300	-0.649	-1.40	-2.15
	AR	-0.684	-1.16	-1.72	0.126	-0.130	-0.339	-	-	-
2024	TV-VAR	-0.325	-0.647	-0.946	-0.004	-0.537	-1.36	-0.318	-1.17	-2.30
	VAR	-0.304	-0.548	-0.766	0.447	0.178	0.029	0.197	-0.271	-0.687
	AR	-0.273	-0.530	-0.739	0.562	0.376	0.300	-	-	-

Table IA.2: **Mean Log Scores.** Average marginal log scores for normalized reserves and the federal funds-IORB spread, together with the average joint log scores, from the TV-VAR (A.8) and the VAR and AR models described in Internet Appendix B.

on $[0, 1]$. Figure IA.2 plots the empirical cumulative distribution functions (CDFs) of the PITs for each horizon and each series across different models. For a well-calibrated forecast, the PIT should have a CDF matching that of a uniform distribution, i.e., a 45-degree line.

For normalized reserves, the empirical CDF of the PIT from the TV-VAR is close to the 45-degree line and within its 90% confidence bands at all horizons; in contrast, the empirical CDFs of the PITs from the VAR and AR models tend to be consistently above the 45-degree line and outside their confidence bands over a sizable share of the $[0, 1]$ support, especially at longer horizons. This is particularly evident for the AR forecasting model at horizon $h = 20$.

For the spread, all models are less well calibrated: the CDFs of the PITs for the spread forecasts are further away from 45-degree line than those for the reserve forecasts, especially at longer horizons. That CDF of the PIT from the TV-VAR, however, has a clear sigmoid shape crossing the 45-degree line from below around 0.5, which suggests that the predictive distribution is quite dispersed but centered around the realized data. The PITs' CDFs from the VAR and AR models, instead, tend to be below the 45-degree line for most of the $[0, 1]$ support, suggesting that the predictive distributions may be biased; this is particularly visible for the VAR model at horizons $h = 10$ and 20.

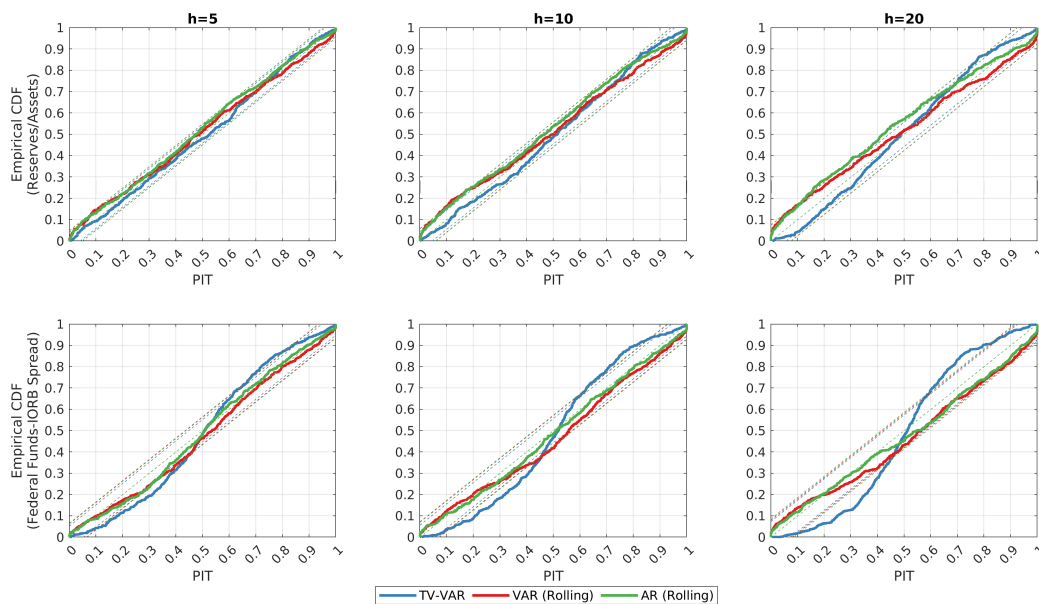


Figure IA.2: **Empirical cumulative distribution function (CDF) of the probability integral transforms (PITs)**. Empirical CDFs of the PITs of the forecasts of normalized reserves and of the federal-funds-IORB spread from the TV-VAR (A.8) and the VAR and AR models described in Internet Appendix B. The colored dashed lines represent the 90% confidence bands for the different models, constructed using the bootstrap method outlined in Rossi and Sekhposyan (2019).

Internet Appendix C Different Normalizations of Reserves

Internet Appendix C.1 Normalization by bank deposits

In this section, we replicate the main results of the paper using aggregate reserves normalized by commercial banks' deposits, instead of reserves normalized by banks' assets. Weekly data on deposits of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("DPSACBW027SBOG"). We linearly interpolate these weekly data to obtain a daily series. Figure IA.3 shows the evolution of reserves normalized by deposits over 2010-2024. Compared to reserves normalized by banks' assets, the level is higher (as bank deposits are smaller than bank assets), but the time variation is almost identical.

Figure IA.4 shows our time-varying IV estimate of the elasticity of the federal funds-IORB spread to shocks in deposit-normalized reserves, using as instrument the forecast errors from the baseline bivariate model (A.8); Figures IA.5, and IA.6 show the robustness results from the trivariate models controlling for repo rates and Treasury yields, respectively. Results are almost identical to those obtained using asset-normalized reserves (see Figures 5, 6, and 7): the elasticity is significantly negative in 2010-2011 and from early 2018 to mid-2020, while being indistinguishable from zero during 2012-2017 and after mid-2020.

Table IA.3 reports the posterior median elasticity to deposit-normalized reserves by year. Quantitatively, the elasticity is smaller in absolute value because the ratio of reserves to deposits is larger than the ratio of reserves to assets. In economic terms, however, results are comparable: a decline in deposit-normalized reserves equal to one standard deviation of the daily changes in reserves (0.27 pp) leads to an increase in the federal funds-IORB spread by 0.24 bp in 2010 and 0.16 bp in 2019, which correspond to 25% and 20% of the standard deviation of the daily changes in the spread. These effects are similar to those estimated using bank assets to normalize reserves (see Table 1).

These results show that our findings are not sensitive to the choice of the normalization factor; the normalization only controls for a time trend in nominal reserves. The variation used to identify the slope of the demand curve in our sample comes from exogenous fluctuations in the supply of reserves (i.e., from the numerator, not from the denominator).

2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
(a) Bi-variate Model														
-0.9 (-1.4,-0.3)	-0.7 (-1.1,-0.2)	-0.3 (-0.8,0.2)	-0.2 (-0.6,0.3)	0 (-0.3,0.3)	0 (-0.3,0.2)	-0.2 (-0.4,0.1)	-0.1 (-0.4,0.1)	-0.5 (-0.8,-0.1)	-0.6 (-0.9,-0.2)	-0.2 (-0.7,0.2)	0 (-0.3,0.2)	-0.1 (-0.3,0.1)	-0.1 (-0.3,0.1)	-0.1 (-0.3,0.2)
(b) Tri-variate Model with Repo Rates														
-1.3 (-1.9,-0.7)	-0.9 (-1.5,-0.4)	-0.7 (-1.2,-0.1)	-0.5 (-1,0.1)	-0.1 (-0.5,0.2)	-0.1 (-0.4,0.2)	-0.2 (-0.4,0.1)	-0.1 (-0.4,0.1)	-0.4 (-0.8,-0.1)	-0.7 (-1.1,-0.3)	-0.2 (-0.8,0.3)	-0.1 (-0.3,0.2)	-0.1 (-0.3,0.2)	-0.1 (-0.3,0.2)	-0.1 (-0.3,0.2)
(c) Tri-variate Model with Treasury Yields														
-0.9 (-0.4,0.3)	-0.7 (-1.5,-0.3)	-0.4 (-1.2,-0.2)	-0.3 (-0.9,0.2)	-0.1 (-0.8,0.2)	0 (-0.4,0.3)	-0.2 (-0.3,0.3)	-0.1 (-0.4,0.1)	-0.5 (-0.4,0.2)	-0.6 (-0.9,-0.1)	-0.2 (-1,-0.2)	0 (-0.7,0.3)	0 (-0.3,0.3)	0 (-0.3,0.3)	-0.1 (-0.3,0.3)

Table IA.3: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves by year.** The estimate of elasticity is obtained as the posterior median of the ratio between the impulse response of the federal funds rate and the impulse response of reserves to a forecast error in reserves at a five-day horizon; see equation (9). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (A.8); in panel (b), they are estimated from an augmented trivariate version of model (A.8) that also includes daily repo rates; in panel (c), they are estimated from an augmented trivariate version of model (A.8) that includes daily Treasury yields. All multivariate models include ten lags ($m = 10$ days). The reported elasticities are calculated by year. Reserves are measured as a ratio to bank deposits, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

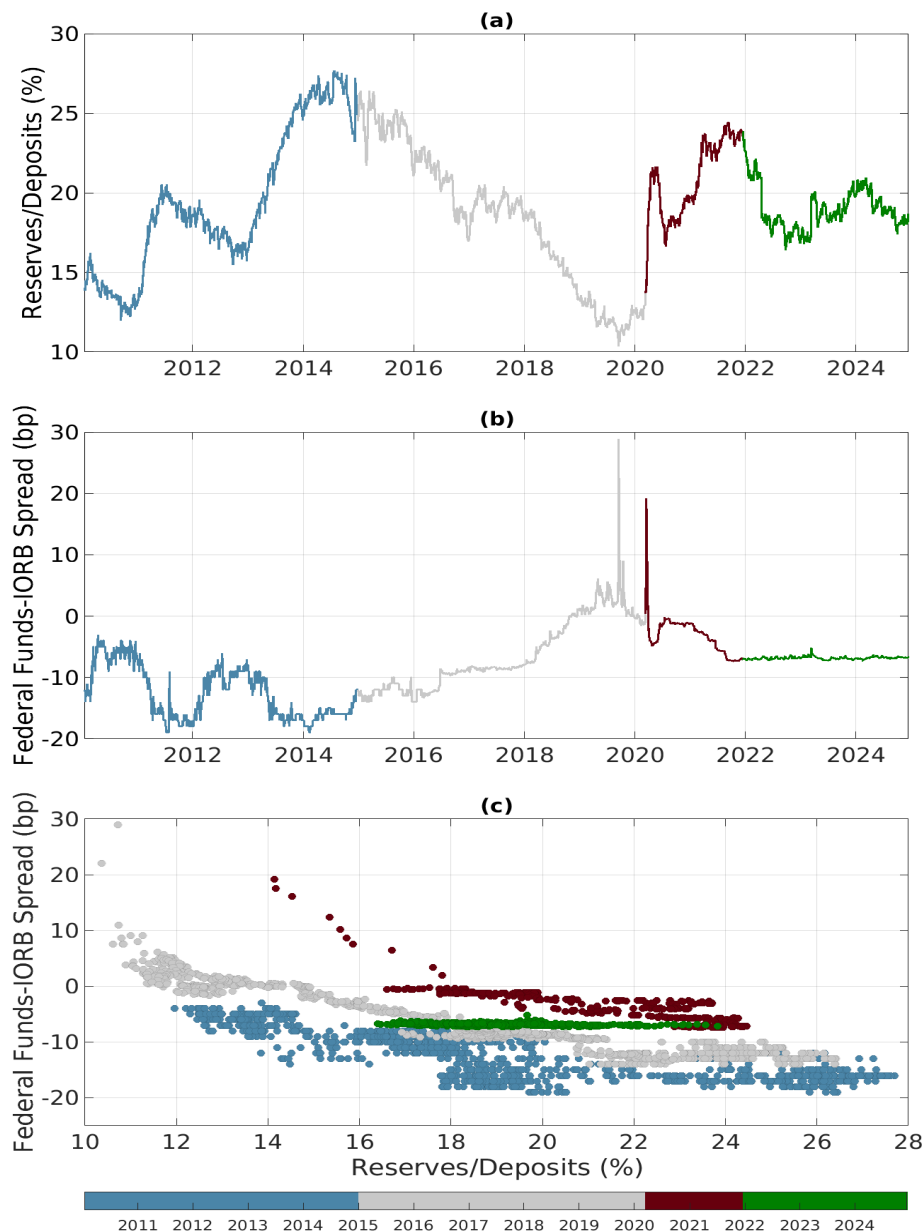


Figure IA.3: **Deposit-normalized reserves, the federal funds rate, and the reserve demand curve.** Panel (a) plots aggregate reserves relative to to commercial banks’ deposits from January 1, 2010 to December 13, 2024. Panel (b) shows the spread between the volume-weighted average federal funds rate and the IOB rate (in basis points). Panel (c) plots the relationship between the spread and normalized reserves.

Internet Appendix C.2 Normalization by GDP

In this section, we replicate the main results of the paper using aggregate reserves normalized by gross domestic product (GDP), instead of reserves normalized by banks’ assets. Quarterly

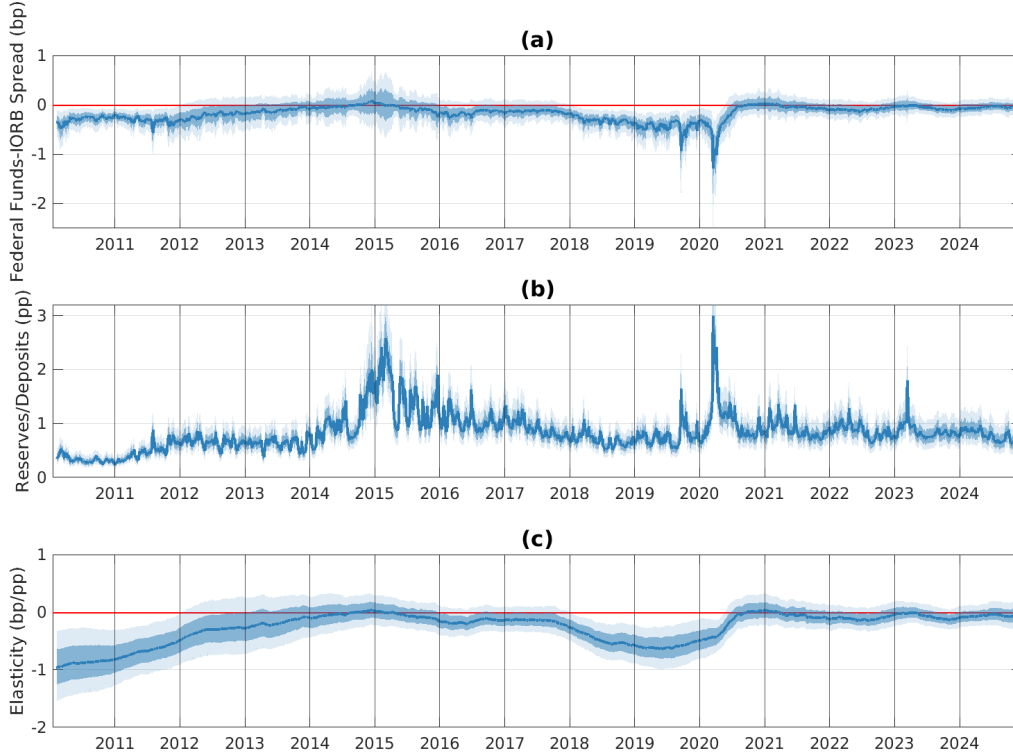


Figure IA.4: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample from model (A.8) with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to bank deposits, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

data on U.S. GDP are from FRED; we linearly interpolate these quarterly observations to obtain daily estimates. Figure IA.7 shows the evolution of reserves normalized by GDP over 2010-2021. Compared to reserves normalized by banks' assets, the level is lower by roughly 2 pp (as GDP is greater than banks' assets), but the time variation is almost identical.

Figure IA.8 shows our time-varying IV estimate of the elasticity of the federal funds-IORB spread to shocks in GDP-normalized reserves, using as instrument for normalized reserves the forecast errors from the baseline bivariate model (A.8); Figures IA.9, and IA.10 show the robustness results from the trivariate models controlling for repo rates and Treasury yields, respectively. Results are almost identical to those obtained using asset-normalized

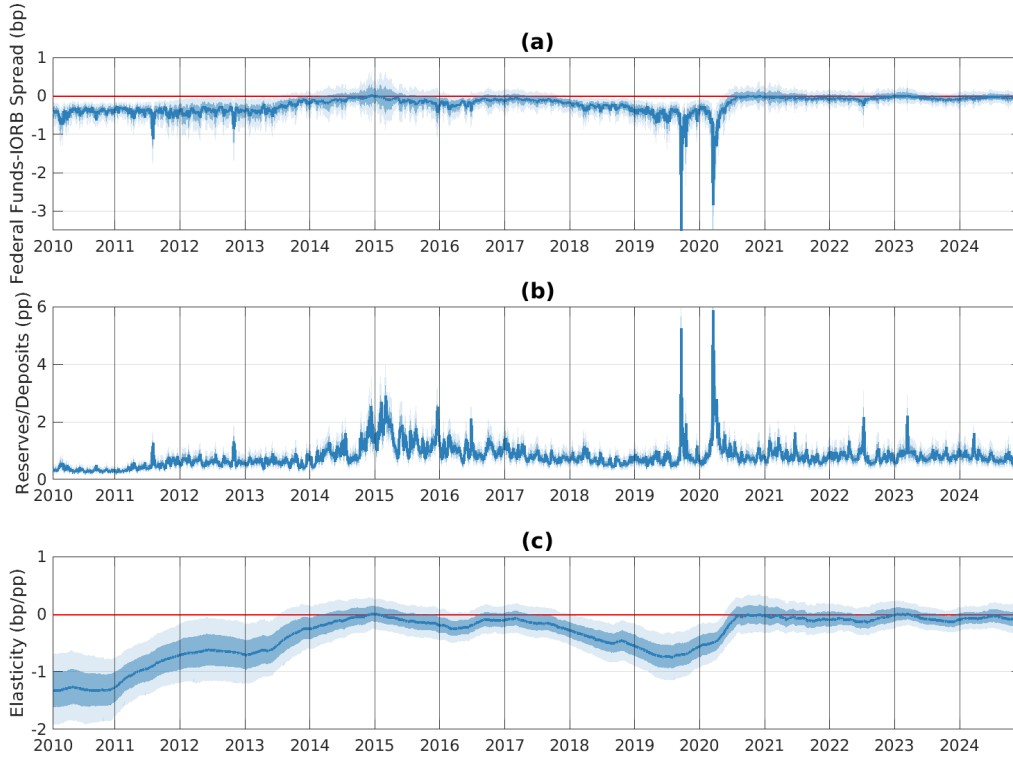


Figure IA.5: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves controlling for repo rates.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (A.8) that includes daily repo rates, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to bank deposits, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

reserves (see Figures 5, 6, and 7): the elasticity is significantly negative in 2010-2011 and from early 2018 to mid-2020, while being indistinguishable from zero during 2012-2017 and after mid-2020.

Results are also quantitatively close to those obtained using asset-normalized reserves. Table IA.4 reports the posterior median elasticity to GDP-normalized reserves by year: a 1-pp drop in reserves normalized by GDP would lead to an increase in the federal funds-IORB spread by 1.7 bp in 2011 and by 1 bp in 2019, while having no effect in 2014 or early 2021. These numbers are close to those reported in Table 1 of the main text.

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
(a) Bi-variate Model															
	-1.7	-1.2	-0.6	-0.3	0	-0.1	-0.2	-0.2	-0.8	-1	-0.3	-0.1	-0.1	-0.1	0
	(-2.8,-0.6)	(-2.1,-0.3)	(-1.5,0.4)	(-1.1,0.6)	(-0.6,0.6)	(-0.5,0.4)	(-0.7,0.2)	(-0.6,0.2)	(-1.4,-0.2)	(-1.6,-0.3)	(-1.1,0.4)	(-0.4,0.4)	(-0.5,0.2)	(-0.4,0.3)	(-0.4,0.4)
(b) Tri-variate Model with Repo Rates															
	-2.5	-1.8	-1.1	-0.8	-0.2	-0.1	-0.3	-0.2	-0.8	-1.2	-0.3	-0.1	-0.1	-0.1	0
	(-3.7,-1.4)	(-2.9,-0.7)	(-2.1,-0.1)	(-1.9,0.2)	(-0.9,0.5)	(-0.7,0.4)	(-0.8,0.2)	(-0.7,0.3)	(-1.4,-0.1)	(-1.9,-0.5)	(-1.3,0.4)	(-0.5,0.4)	(-0.5,0.3)	(-0.5,0.3)	(-0.4,0.4)
(c) Tri-variate Model with Treasury Yields															
	-1.8	-1.4	-0.7	-0.5	-0.1	-0.1	-0.3	-0.2	-0.9	-1.2	-0.3	-0.1	0	0	-0.1
	(-3,-0.7)	(-2.3,-0.3)	(-1.7,0.3)	(-1.4,0.5)	(-0.8,0.6)	(-0.6,0.5)	(-0.8,0.2)	(-0.7,0.3)	(-1.6,-0.2)	(-1.9,-0.5)	(-1.2,0.5)	(-0.5,0.4)	(-0.5,0.5)	(-0.5,0.5)	(-0.5,0.5)

Table IA.4: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves by year.** The estimate of elasticity is obtained as the posterior median of the ratio between the impulse response of the federal funds rate and the impulse response of reserves to a forecast error in reserves at a five-day horizon; see equation (9). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (A.8); in panel (b), they are estimated from an augmented trivariate version of model (A.8) that also includes daily repo rates; in panel (c), they are estimated from an augmented trivariate version of model (A.8) that includes daily Treasury yields. All multivariate models include ten lags ($m = 10$ days). The reported elasticities are calculated by year. Reserves are measured as a ratio to GDP, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

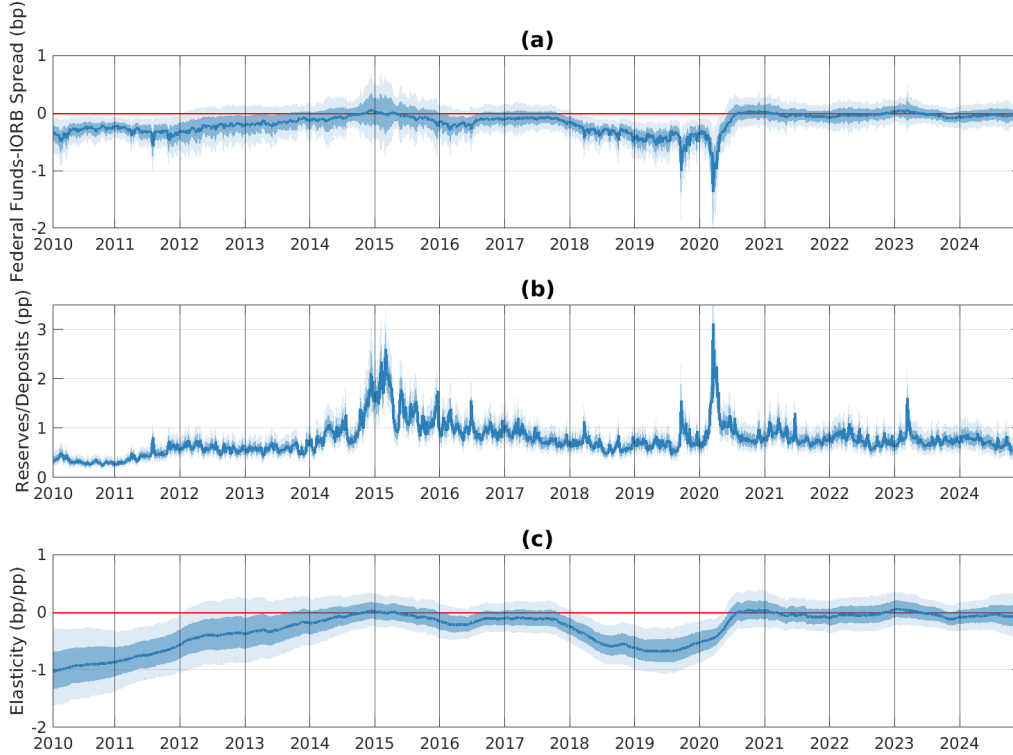


Figure IA.6: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves controlling for Treasury yields.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (A.8) that includes daily yields on 1-year U.S. Treasury securities, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to bank deposits, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

Internet Appendix D Sign-Restricted Time-Varying VARs

In this appendix, we replicate the main results of the paper using a structural time-varying VAR identified with sign restrictions. We identify daily demand and supply disturbances using sign restrictions (Uhlig, 2005): we assume that demand shocks induce positive comovement between reserves and spreads, while supply shocks induce negative comovement. We assume that these restrictions hold on impact and for the following five days. Panel (c) of Figure IA.11 shows the time-varying estimates of the elasticity of the federal funds-IORB spread

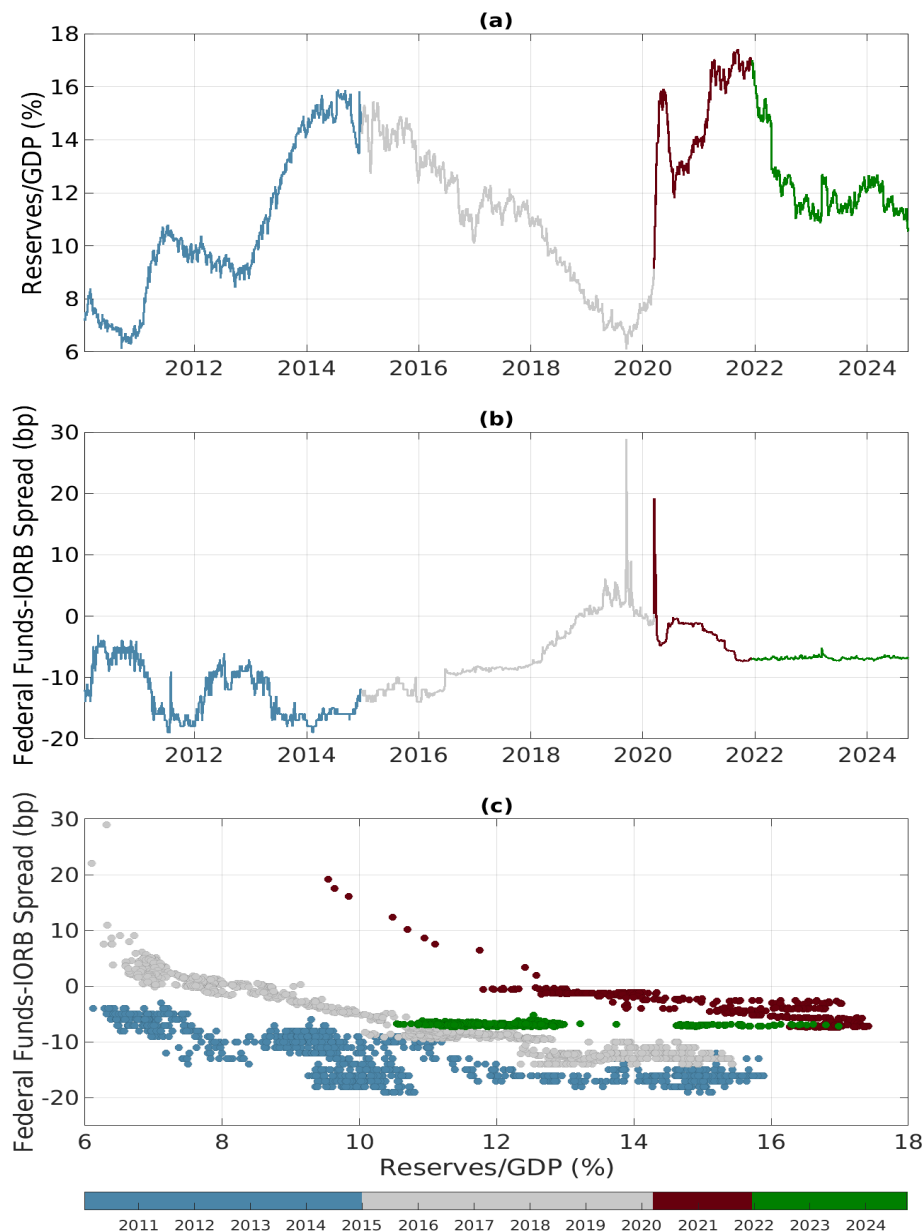


Figure IA.7: **GDP-normalized reserves, the federal funds rate, and the reserve demand curve.** Panel (a) plots aggregate reserves relative to GDP from January 1, 2010 to December 13, 2024. Panel (b) shows the spread between the volume-weighted average federal funds rate and the IOB rate (in basis points). Panel (c) plots the relationship between the spread and normalized reserves.

to shocks in reserves (normalized by commercial banks' assets) from the sign-restricted VAR. Results are similar to our estimates using the baseline model (A.8), albeit more volatile.

To improve identification and mitigate known pathologies of pure sign-restriction approaches (Baumeister and Hamilton, 2015), we impose an additional inequality restriction

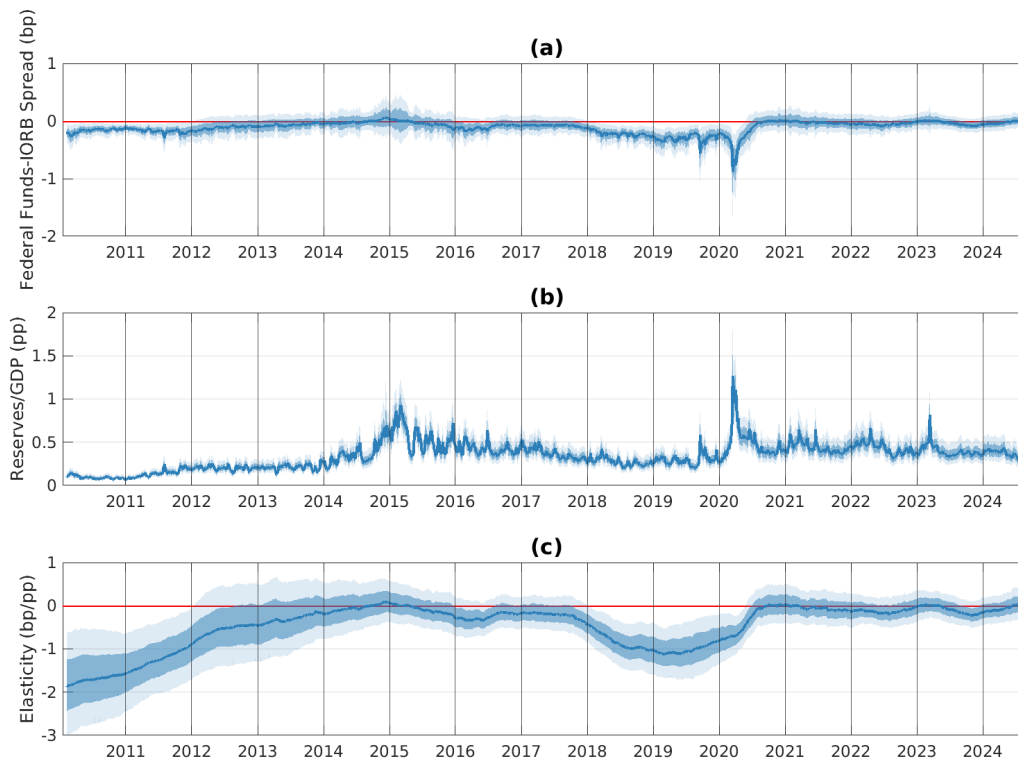


Figure IA.8: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample from model (A.8) with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to GDP, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

that limits the same-day response of reserves to demand shocks, consistent with the current monetary policy implementation framework (Section 2.2). Namely, we impose that the contemporaneous response of reserves to a one-basis-point demand shock in the federal funds-IORB spread must be smaller than 0.01 percentage points of bank assets. This upper bound is consistent with the Federal Reserve injection of reserves following the turmoil of September 2019: the spread increased by 25 bp over September 16 and 17, and the Federal Reserve operations only started on September 17 with a take-up of \$41 billion, i.e., 0.23% of bank assets at that time. As shown in Figure IA.12, when adding this bound, the elasticity is smoother and closer in magnitude to that of our baseline VAR (A.8).

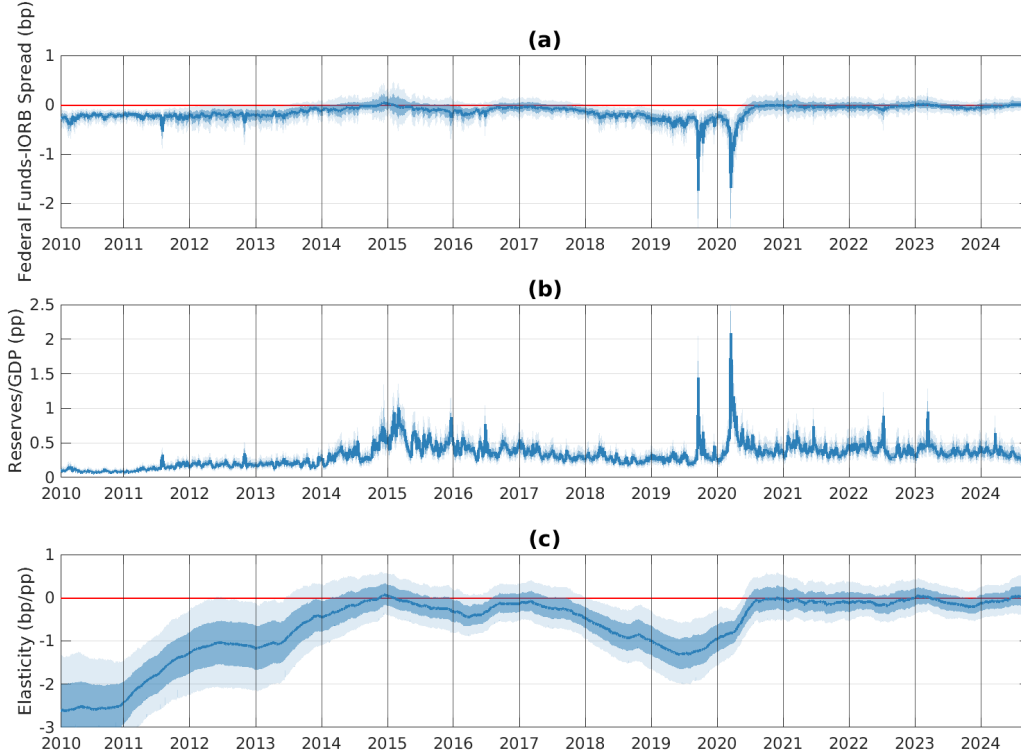


Figure IA.9: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves controlling for repo rates.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (A.8) that includes daily repo rates, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to GDP, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

To estimate these structural time-varying VARs with inequality restrictions, we use a two-step approach. First, we estimate the reduced-from model; second, following Rubio-Ramírez et al. (2010), we apply 20,000 random rotation matrices to the impulse response functions defined in Appendix B.1 and retain only those satisfying all restriction. This two-step approach is convenient, as it keeps the estimation of the reduced-form model separate from the identification of structural shocks. While we follow this approach here, recent advances make it possible to perform estimation and identification jointly in a single step, which may improve efficiency (Arias et al., 2025).

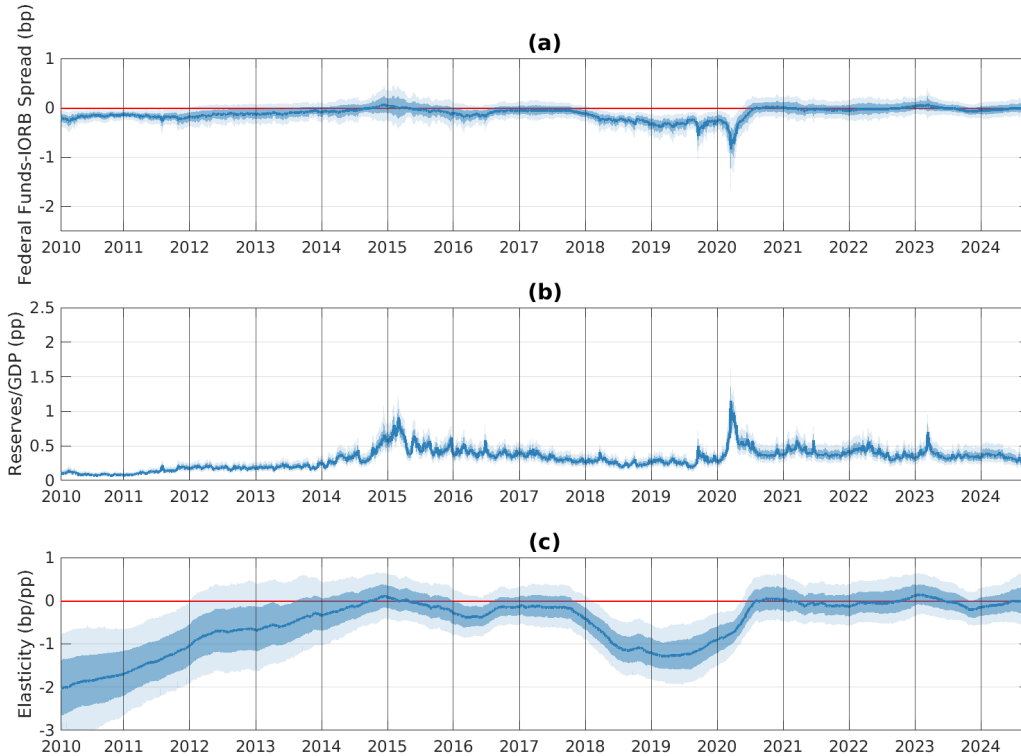


Figure IA.10: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves controlling for Treasury yields.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (A.8) that includes daily yields on 1-year U.S. Treasury securities, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to GDP, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

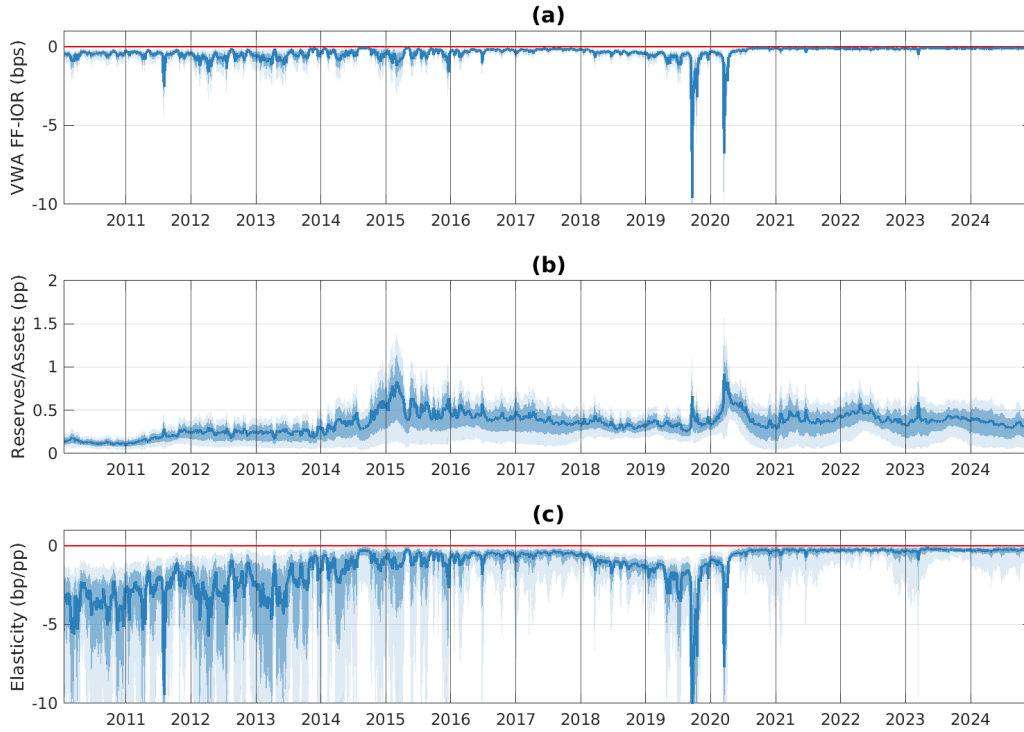


Figure IA.11: **IV estimate of the elasticity of the federal funds rate to reserves normalized by bank assets using sign-restricted VAR.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a sign-restricted time-varying VAR, with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to GDP, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

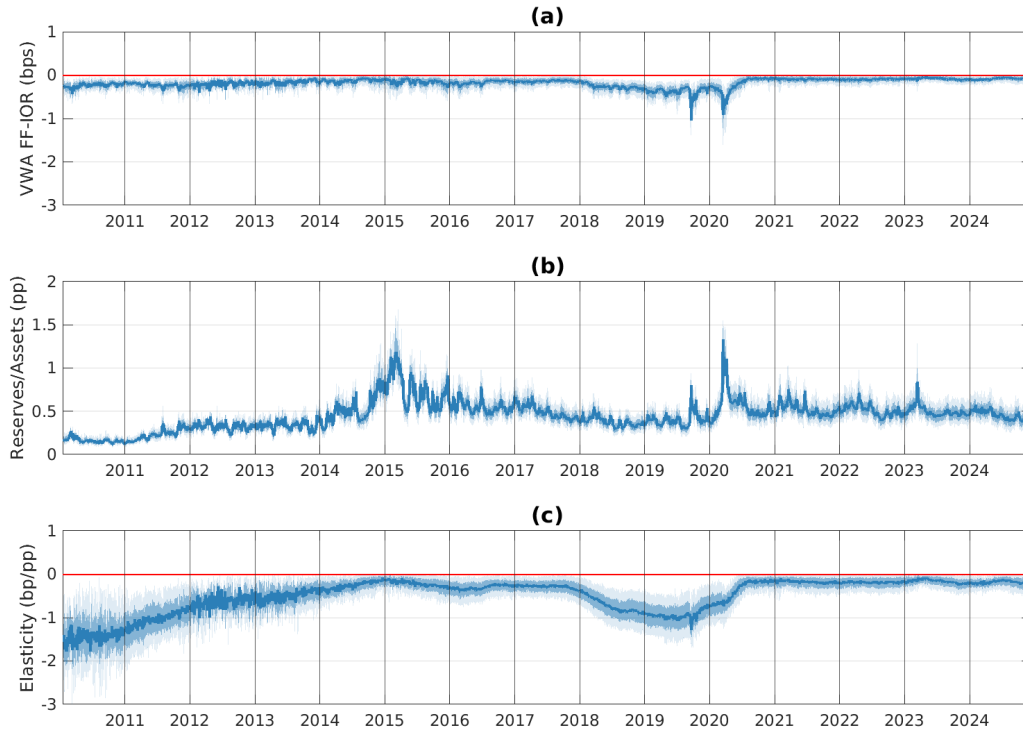


Figure IA.12: **IV estimate of the elasticity of the federal funds rate to reserves normalized by bank assets using sign-restricted VAR with zero contemporaneous response of supply to demand shocks.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (9). Forecast errors and impulse responses are estimated in-sample using a sign-restricted time-varying VAR with ten lags ($m = 10$ days) that imposes no same-day reserve supply response to demand shocks. The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to the 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to GDP, in percent. The federal funds rate is measured as a spread to the IORB rate, in basis points.

Internet Appendix E Indicator of Reserve Ampleness Based on Quarter-End Repo Rates

An alternative indicator to the daily share of overnight repo transacted at rates above the IORB is the size of the quarter-end spikes in repo rates. As we discuss in Section 2.3, banks have incentives to window dress and contract their balance sheets on regulatory reporting dates to report more favorable regulatory metrics (Bassi et al., 2024). On quarter-ends, the balance-sheet costs of European banks and their affiliated dealers increase, leading to a decline in repo intermediation and higher repo rates. To keep repo rates close to the IORB on those reporting dates, U.S. banks must be willing to increase their repo lending to accommodate the drop in intermediation by European banks. The size of the repo spikes thus depends on the level of aggregate reserves: as reserves become less ample, U.S. banks will be more constrained, and quarter-end repo rates increase more (Correa et al., 2020; Copeland et al., 2025).

Figure IA.13 shows the time series of the spread between repo rates and the IORB on quarter ends, along with our OOS estimates of the elasticity, from 2015 to 2024. Consistent with the repo indicator based on the daily share of repo transacted at rates above the IORB (see Section 5.2.3) and with the elasticity estimates, repo spikes are low in 2015-2017, increase in 2018-2019, remain low since 2020. Figure Internet Appendix E shows a version of the spider-web chart in Figure 11 that includes the size of quarter-end repo spikes as an indicator of reserve ampleness.

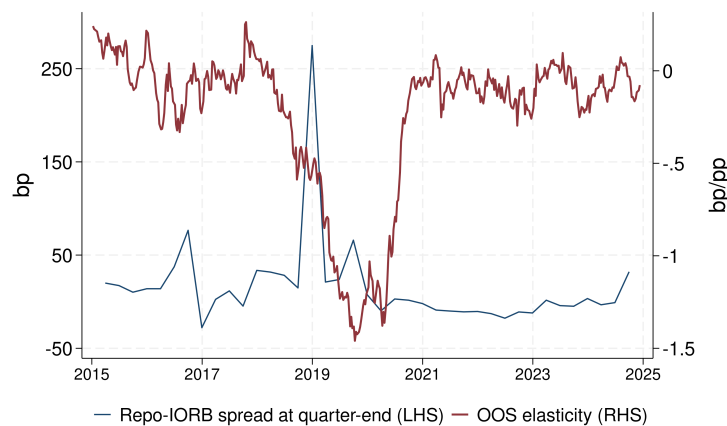


Figure IA.13: **Indicator of reserve ampleness: repo rates quarter-end spikes.** The blue line shows the size of spikes in repo rates relative to the IORB on quarter ends from January 1, 2015 to December 13, 2024.

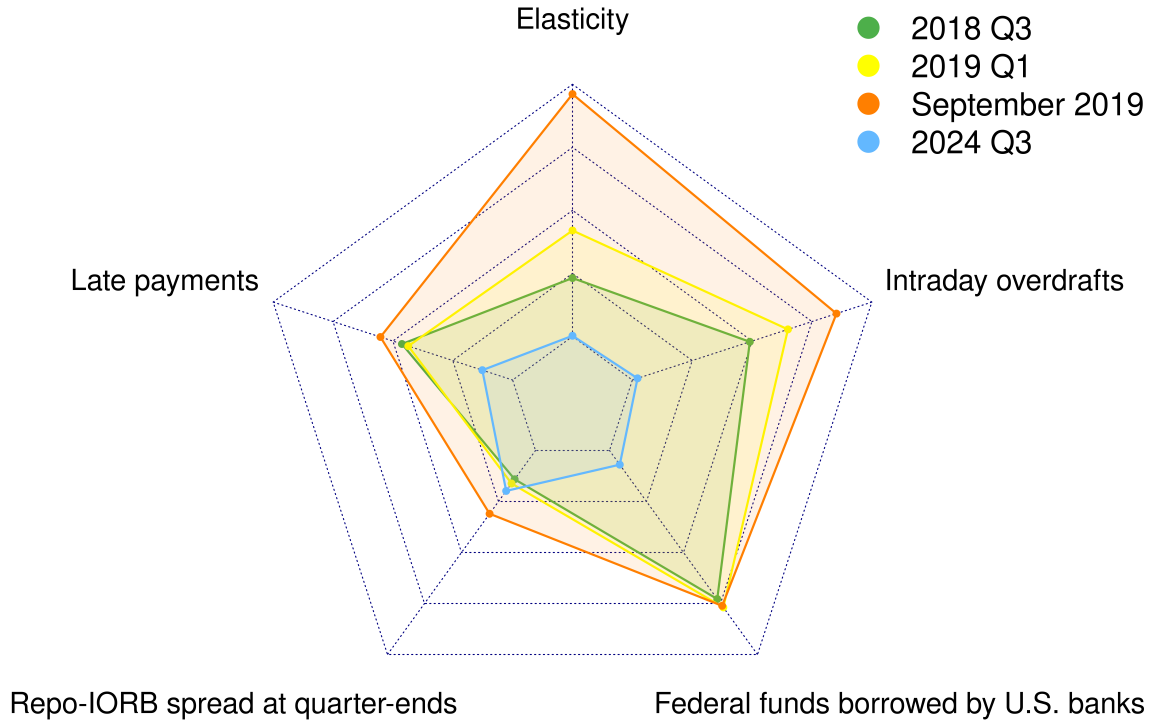


Figure IA.14: **Indicators of reserve amplexness: spider web chart.** This chart plots the OOS estimates of the reserve demand elasticity and four additional indicators of reserve amplexness in four different periods. The additional indicators are: (i) size of quarter-end repo rates spikes (relative to the IORB). Data on interdealer repo rates and volumes are from the Depository Trust & Clearing Corporation; (ii) the daily share of Fedwire Funds Service payments settled after 5 pm ET (“late payments”); (iii) the banking system average daylight (intraday) overdraft. Account balance data are from the Daylight Overdraft Reporting and Pricing System; (iv) the daily share of federal funds borrowed by U.S. banks. Transactions-level data are collected in the Federal Reserve’s Report of Selected Money Market Rates (FR 2420). Each variable is normalized such that the inner and outer pentagons correspond to the highest and lowest levels of reserve amplexness between 2015 and 2024, as measured by that variable. The four periods are: 2018Q3, 2019Q1, September 2019, and 2024Q4; for each period, we show the average of the daily value of each indicator in that period.

Internet Appendix F Vertical Shifts in Reserve Demand: Additional Evidence

Internet Appendix F.1 2018-2019 changes in the ONRRP-IORB spread

In this appendix, we replicate the analysis of the effect of ONRRP-IORB spread adjustments on the vertical location of the reserve demand curve (Section 6.1), including the adjustments

of 2018–2019. Based on regression (10), we run the following daily regression from January 2014 to May 2019:

$$\Delta p_{t+h,t-1} = \alpha + \beta * \Delta \text{ONRRP-IORB}_{t,t-1} + \beta_{t \geq 2018} * \mathbf{1}[t \in [2018, 2019]] \times \Delta \text{ONRRP-IORB}_{t,t-1} \\ x^c + \gamma_1 * \Delta \text{Controls}_{t+h,t-1} + \gamma_2 * \text{Calendar Controls}_{t+h} + \varepsilon_{t+h}, \quad (\text{IA.1})$$

where the effect of the 2018–2019 ONRRP-IORB adjustments is $\beta + \beta_{t \geq 2018}$, and $\beta_{t \geq 2018}$ is the difference relative to the effect of the 2013–2015 adjustments. All the other variables are defined as in regression (10). Finally, as in Section 6.1.1, we use $h = 0$ and $h = 1$; that is, we look at one-day and two-day changes in the federal funds-IORB spread around changes in the ONRRP-IORB spread.

Results are in Table IA.5; standard errors are Newey-West with five lags. In 2018–2019, the pass-through of ONRRP-IORB adjustments to the federal funds-IORB spread is significantly smaller than in 2013–2015 and statistically insignificant; e.g., the one-day pass-through is only 16% (p -value = 0.18). In Columns (3) and (4), we re-estimate the regression starting the sample in September 2013, to also include the ONRRP-IORB adjustments that were implemented in the early months of the facility existence; results are similar.

	(1)	(2)	(3)	(4)
	$\Delta \text{FFR-IOR}_{t,t-1}$	$\Delta \text{FFR-IOR}_{t+1,t-1}$	$\Delta \text{FFR-IOR}_{t,t-1}$	$\Delta \text{FFR-IOR}_{t+1,t-1}$
$\Delta \text{ONRRP-IOR}_{t,t-1}$	0.369*** (3.71)	0.359*** (4.07)	0.323*** (3.23)	0.311*** (3.46)
$(t \in [2018, 2019]) = 1 \times \Delta \text{ONRRP-IOR}_{t,t-1}$	-0.211 (-1.37)	-0.194 (-1.46)	-0.167 (-1.08)	-0.151 (-1.13)
$\beta + \beta_{t \geq 2018}$	0.158	0.164*	0.156	0.160
p	0.182	0.094	0.190	0.101
R^2	0.128	0.139	0.110	0.124
Sample	1/2014 - 5/2019	1/2014 - 5/2019	9/2013 - 5/2019	9/2013 - 5/2019
Observations	1121	1115	1175	1170

Table IA.5: **One-day and two-day impact of ONRRP-IORB spread adjustments on the federal funds-IORB spread from January 2014 to May 2019 and from September 2013 to May 2019.**

We also replicate the analysis at the monthly horizon. We run regression (11) on the ONRRP-IORB adjustments of June 14, 2018, December 20, 2018, and May 2, 2019. Figure IA.15 shows the results of the regressions estimated on two-month periods around these event (20 business days before and after). Consistent with the daily frequency analysis, the effect of the 2018–2019 ONRRP-IORB spread adjustments on the federal funds-IORB spread is smaller in magnitude than the effect of the 2014–2015 ones. At the one-week horizon, the

pass-through is between 0.15 bp and 0.40 bp; at the one-month horizon, it is between 0.20 bp and 0.52 bp.

As we discuss in Section 6, these results should be taken with a grain of salt because reserves were not abundant in 2018–2019; that is, the federal funds market was operating in the sloped region of the reserve demand curve. As a result, changes in the federal funds-IORB spread may also reflect changes in reserve supply or horizontal shifts in the demand curve. In contrast, the results in Section 6 are for the ONRRP-IORB spread adjustments of 2014–2015, when reserves were abundant; that is, the market was operating in the flat region of the demand curve, where neither supply shocks nor horizontal demand shift have an effect on the federal funds-IORB spread.

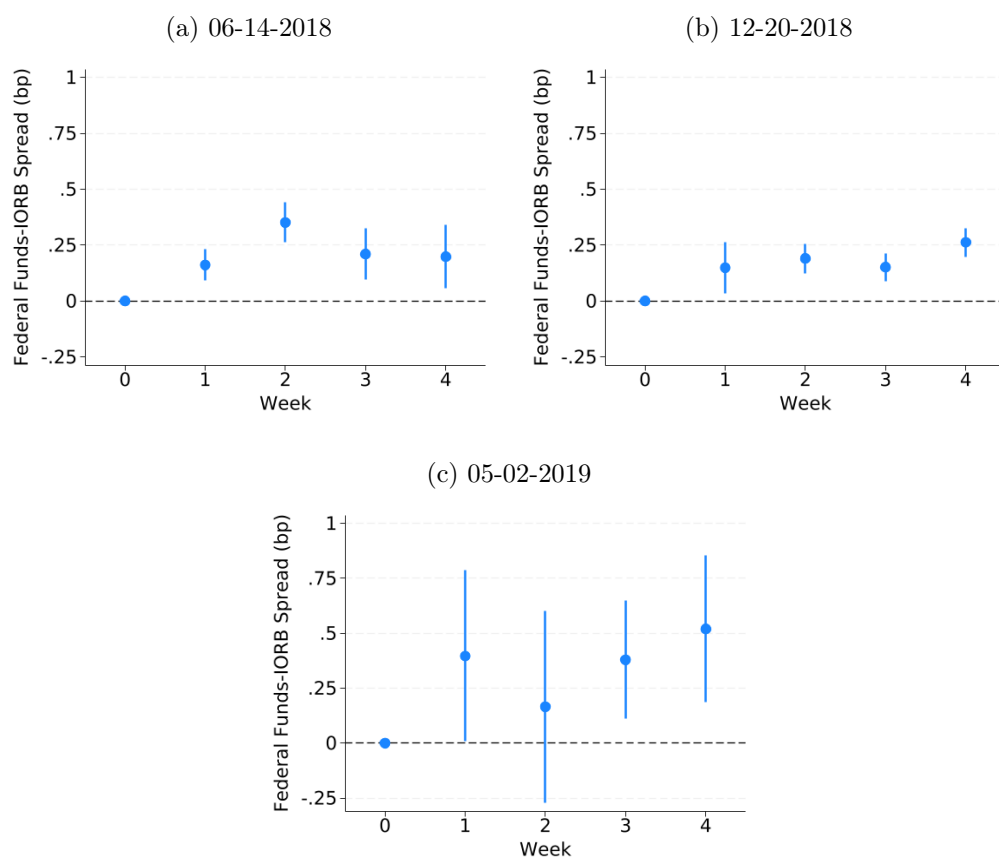


Figure IA.15: Monthly horizon impact of 2018-2019 ONRRP-IORB spread adjustments on the federal funds-IORB spread.

Internet Appendix F.2 2011 Change of the FDIC Assessment Fee

In 2011, the Federal Deposit Insurance Corporation (FDIC) changed the calculation of its assessment fee: instead of paying fees proportional only to their deposit liabilities, banks had to pay fees proportional to their total assets (net of tangible equity). By expanding the base for the fee calculation to the size of banks' balance sheets, this change increased banks' balance-sheet costs. The rule was announced on February 9 and became effective on April 1, 2011.

The model in Appendix A predicts that an increase in banks' balance-sheet costs should decrease the rate at which banks are willing to borrow, pushing the federal funds-IORB spread down. Consistent with theory, Figure IA.16 shows that the federal funds-IORB spread declines around the change of the FDIC assessment fee, with the decline being particularly notable around the implementation date.



Figure IA.16: **The federal funds-IORB spread around the announcement and implementation of the 2011 FDIC change of the assessment fee.** The blue line represents the realized spread. The two dashed vertical lines represent the announcement (2/9/2011) and implementation (4/1/2011) of the change.

To quantify the downward vertical shift caused by the 2011 change in the FDIC assessment fee, we estimate the following daily regression, both over the two months surrounding its announcement and over the two months surrounding its implementation:

$$p_t = \alpha + \sum_{i=1}^4 \beta_i * \mathbf{1}[\text{Post FDIC-Fee Change Week}_i]_t + \gamma * \text{Controls}_t + \varepsilon_t, \quad (\text{IA.2})$$

where $\mathbf{1}[\text{FDIC fee week}_i]_t$ is a dummy for the i -th week after the event, and all other variables are defined as in regression (11). Results are in Figure IA.17, with 90% confidence intervals based on Newey-West standard errors with five lags; panel (a) is for the actual implementation date and panel (b) is for the announcement date.

In the first week after the actual change in the assessment fee, the federal fund-IORB spread declines by 3 bp (p -value < 0.01) relative to its average level in the previous month. The spread does not change significantly in the following weeks; one month after the change in the FDIC assessment fee, the decline relative to the previous month is about 4 bp (p -value < 0.01). The effect around the announcement is significant but smaller, with a relative decline of 1 bp in the first week (p -value = 0.06) and 2 bp over the four weeks after the event (p -value = 0.03).

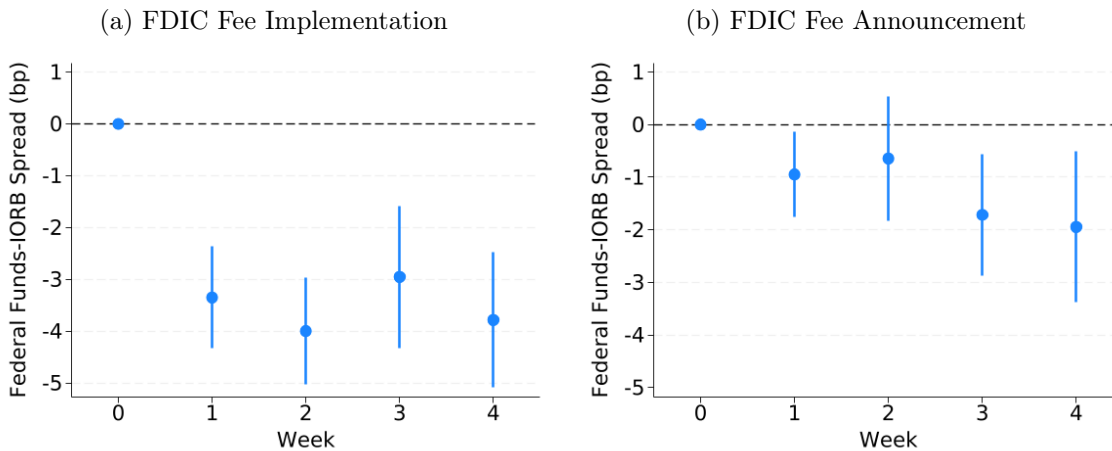


Figure IA.17: **One-month horizon impact of the 2011 FDIC fee assessment change on the federal funds-IORB spread.**

To quantify these effects at longer horizons, we estimate regression (IA.2) expanding the sample to two months before and two months after each event. Results are in Figure IA.18. The impact of the implementation of the 2011 FDIC assessment fee is practically unchanged, whereas the announcement does not seem to significantly affect the federal funds-IORB spread relative to its average level in the two months before the announcement. (The coefficient for the eighth week after the announcements is significantly negative because that week corresponds to the first days after the change in the assessment fee became effective.)

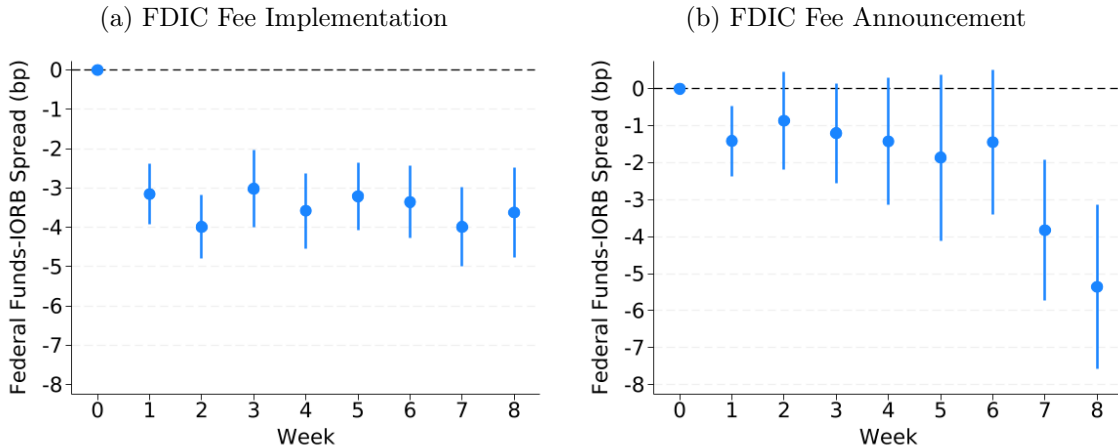


Figure IA.18: **Two-month horizon impact of the 2011 FDIC fee assessment change on the federal funds-IORB spread.**

As the results on the effect of the 2018–2019 ONRRP-IORB spread adjustments, the results on the effect of the 2011 change in the FDIC assessment fee on the federal funds-IORB spread should be taken with caution because reserves were scarce in 2011; therefore, the effect we measure may be due to supply shocks or horizontal demand shifts, rather than to a vertical upward shift caused by an increase in balance-sheet costs. The analysis of the effect of the SLR relief of 2020–2021, in contrast, is robust to this confounding factors because reserves were abundant during that period (Section 6.2).

Internet Appendix G Non-linear Fit of the Reserve Demand Curve

Our estimation methodology is highly flexible and able to identify the time-varying elasticity of the federal funds rate to reserve shocks, but it does not directly allow for the recovery of the reserve demand function. In this section, we develop and implement a method to recover the underlying demand function based on the joint forecasts of prices and quantities from our time-varying VAR model. We assume a nonlinear functional form for the demand curve consistent with the theoretical framework in Section 2.1. We assume that the shape of the demand function is time-invariant but allow for slow-moving structural shifts in its vertical and horizontal locations. In addition to providing an empirical description of the demand for reserves, this exercise also facilitates the assessment of the relative scarcity of the supply of reserves.

Internet Appendix G.1 Post-processing of model forecasts

While our time-varying IV estimates of the slope of the reserve demand curve inform us on the satiation level as reserves transition from ample to abundant, they do not answer two important questions. The first one is whether the demand curve has moved vertically; our IV estimates cannot answer this question because vertical shifts do not affect the curve's slope. Moreover, we cannot use the time-varying intercept from the approximate linear model (5) since that coefficient changes not only if the underlying nonlinear curve moves vertically, but also if we move along the demand curve due to supply shocks or if the curve moves horizontally due to low-frequency structural changes.

Knowing whether structural factors move the reserve demand curve up or down is important for several reasons. Structural vertical shifts can permanently push the federal funds rate closer to the bounds of its target range, increasing the probability that the policy rate moves outside its range. Moreover, if persistent vertical shifts are present, there is no one-to-one mapping between the federal funds-IORB spread and the slope of the curve, which means that the spread cannot be used as a proxy for the rate elasticity to reserve shocks. After 2008, in fact, one may be tempted to use the federal funds-IORB spread to make inference on the curve's slope because, according to the theory, both the curve and the absolute value of its slope are strictly decreasing in the region between scarce and abundant reserves (see Section 2.1).⁴⁵ As a result, absent vertical shifts, an increase in the spread would imply an increase in the rate elasticity, suggesting that reserves are becoming scarcer. In the presence of vertical shifts, however, this one-to-one mapping no longer holds, and trends in the federal funds-IORB spread do not necessarily reflect changes in the rate elasticity as is often assumed.

The second question that our locally-linear IV estimates of the slope do not address directly is the transition between ample and scarce reserves. Based on the theory, the distinction between abundant and ample reserves is clear: the minimum level of reserves above which the demand curve is flat; that is, the point of demand satiation. In the empirical analysis, this definition naturally maps into the level of reserves above which the rate elasticity to reserve shocks is statistically insignificant at a given confidence level.

The distinction between ample and scarce reserves, in contrast, is more arbitrary: loosely speaking, reserves transition from ample to scarce when the slope goes from gently negative to very steep. A natural way to formalize this statement is by looking at the rate at which

⁴⁵In the left part of the demand curve, instead, the absolute slope increases with reserves because the curve must flatten around the DW rate as reserves go to zero (Figure 1).

the absolute slope increases as reserves decrease (i.e., the curve’s second derivative): we could define the transition between ample and scarce as the reserve level for which this rate reaches its maximum. To operationalize this definition, however, we need a nonlinear model.

Consistent with the reserve demand curve implied by the theory in equation (4), we specify the following demand function for reserves:

$$p_t = p_t^* + f(q_t - q_t^*; \theta) \quad \text{with} \quad f(x; \theta) = \left(\arctan \left(\frac{\theta_1 - x}{\theta_2} \right) + \frac{\pi}{2} \right) \theta_3, \quad (\text{IA.3})$$

where p^* and q^* are the vertical and horizontal locations, and $\theta = (\theta_1, \theta_2, \theta_3)$ is a vector of parameters characterizing the shape of the curve: θ_1 is a location parameter, θ_2 is a scale parameter, and θ_3 is a normalization factor. We choose this transformation of the arctan function because it has a smooth and decreasing sigmoid shape that goes to zero as $x \rightarrow \infty$, as predicted by the theory (see equation (3)). Consistent with the evidence in Figure 3, we consider three periods, corresponding to different locations of the curve: 2010–2014, 2015–3/09/2020, and 3/16/2020–12/15/2021. We assume that q^* and p^* change across periods but do not change within each period (e.g., $q_t^* = q_1^*$ and $p_t^* = p_1^*$ for all t in 2010-2014), whereas θ is constant across periods.

We do not include the latest part of our sample, 2022–2024, because the elasticity was zero throughout that time; that is, there are no points on the sloped part of the demand curve, which prevents us from identifying a horizontal shift (any horizontal shift would be consistent with those data).

Our post-processing exercise finds the parameters $\{\theta_1, \theta_2, \theta_3, (p_1^*, q_1^*), (p_2^*, q_2^*), (p_3^*, q_3^*)\}$ that minimize the following objective function:

$$\sum_{k=1}^3 \sum_{t \in T_k} \sum_{i=1}^N [p_{it} - p_k^* - f(q_{it} - q_k^*; \theta)]^2 \quad (\text{IA.4})$$

where T_1 , T_2 , and T_3 represent 2010-2014, 2015-3/09/2020, and 3/16/2020-12/29/2021; $i = 1, \dots, N$ are draws from the in-sample five-day-ahead joint posterior distribution of the federal funds-IORB spread (p) and normalized reserves (q) from our bivariate time-varying VAR model (A.8).⁴⁶ We generate these forecasts every five days and set $N = 100$. To improve efficiency and reliability, we also provide the optimization algorithm with the analytical gradient of the objective function (IA.4).

In other words, we perform a nonlinear least-square fit on the time-varying joint forecasts

⁴⁶We choose five-day-ahead forecasts to be consistent with our instrument in the IV analysis.

of prices and quantities from our forecasting model; in this way, we can exploit an entire cross-section of pseudo-data at each point in time, as opposed to one single observation as in the realized times series. This approach leverages the forecasting accuracy of our time-varying VAR model, but in contrast to our IV estimates of the rate elasticity, its results cannot be interpreted causally.

Internet Appendix G.2 Parameter interpretation and constraints

Our estimation method is silent regarding the origins of the vertical and horizontal shifts in (IA.3). Based on the economic theory and institutional framework discussed in Section 2.1, several factors can lead to structural shifts in the demand for reserves. For example, p^* denotes the lower asymptote of the demand curve, which represents the (negative) wedge between the federal funds and IORB rates when reserves are abundant; any factors affecting banks' ability to run the IORB arbitrage, such as balance-sheet costs or the bargaining power of FHLBs (the main federal funds lenders), affect this wedge. Horizontal shifts in q^* , in contrast, reflect factors that shift the demand for reserves at every price level, including changes in liquidity regulation and supervision as well as banks' response to such changes. For normalization, we set $q_1^* = 0$ and interpret q_2^* and q_3^* as horizontal shifts of the curve relative to its 2010-2014 position.

Regarding the time-invariant nonlinear part of the demand curve, θ_1 represents the point of maximum absolute slope, i.e., the reserve level at which the negative slope of the curve is the steepest. We can think of the region around θ_1 as the region of scarce reserves, where the federal funds rate is highly sensitive to even small reserve shocks. The point of maximum slope growth, instead, is $x = \theta_1 + \theta_2/\sqrt{3}$; this point is where the curve's absolute slope increases at the highest rate as reserves decrease, which we interpret as the threshold between ample and scarce reserves.⁴⁷ θ_3 measures the vertical distance between the upper and lower asymptotes of the nonlinear time-invariant function in (IA.3): $\lim_{x \rightarrow -\infty} f(x; \theta) - \lim_{x \rightarrow +\infty} f(x; \theta) = \pi\theta_3$. The theory predicts that, as reserves decline, the federal funds rate should converge (from below) to the DW rate plus a spread capturing balance-sheet costs and other frictions.⁴⁸ As a result, θ_3 should be (at least) of the same order of magnitude as the DW-IORB spread.

To ensure that minimizing (IA.4) leads to economically meaningful results, we use the

⁴⁷Since the arctan function is strictly decreasing, there is no reserve range where curve (IA.3) is perfectly flat, which would correspond to the abundant-reserve region. For inference on the transition between abundant and ample reserves, we rely on the IV estimates of the rate elasticity in Section 5, which suggest that this transition occurs for reserves between 12% and 13% of bank assets, depending on the time period.

⁴⁸Stigma or borrowing caps may also push the federal funds rate above the DW rate.

framework of Section 2.1 to initialize the parameters and set bounds on them. Internet Appendix G provides a detailed description of the algorithm, parameter bounds, and initialization values. Importantly, none of our parameter estimates are equal to their bounds or initial values.

Finally, our IV estimates show that, below a given reserve threshold, the slope of the demand curve becomes increasingly negative as reserves decrease. This evidence indicates that, in our sample, the federal funds market has operated to the right of the scarcity region. For this reason, we minimize (IA.4) imposing the constraint that the algorithm only fits the right tail of the curve (i.e., $q_{it} - q_t^* > \theta_1$ for all t); in robustness checks, we use milder constraints and obtain similar results.

Internet Appendix G.3 Results

Figure IA.19 shows the results of our nonlinear least-squares fit, with low-frequency horizontal and vertical shifts, evaluated on the joint forecasts of prices and quantities from the time-varying VAR (A.8). The estimates of the shifts are in Table IA.6; the estimates of the parameters governing the nonlinear shape of the curve are in Table IA.7.

The reserve demand curve has moved vertically and horizontally over time. As shown in Table IA.6, from 2010–2014 to 2015–2019, it moved upward by roughly 2 bp and to the right by roughly 3 pp. This horizontal shift is consistent with our IV estimates of the rate elasticity, which suggest that the reserve level at which the curve starts displaying a significantly negative slope was higher in the second part of the sample. In 2020–2021, the horizontal location of the curve does not seem to change; the vertical location, in contrast, jumps further up by additional 8 bp, for a total increase of 10 bp relative to the first period (2010–2014). These shifts, and especially the vertical ones, are economically material, as the in-sample standard deviation of the federal funds-IORB spread is around 6 bp and that of normalized reserves is around 3 pp.

These results confirm the evidence in Figure 3 and suggest that, although there seems to be a horizontal shift to the right over 2010–2021, upward vertical shifts seem to be the more relevant source of time variation in the reserve demand curve, especially in the last part of the sample. This result is particularly important because, as we discuss above, the presence of vertical shifts implies that a rise in the federal funds rate relative to the IORB rate cannot be interpreted as a signal of increased reserve scarcity. To identify the transition between abundant and scarce reserves, instead, we need to directly estimate the rate elasticity to reserve shocks.

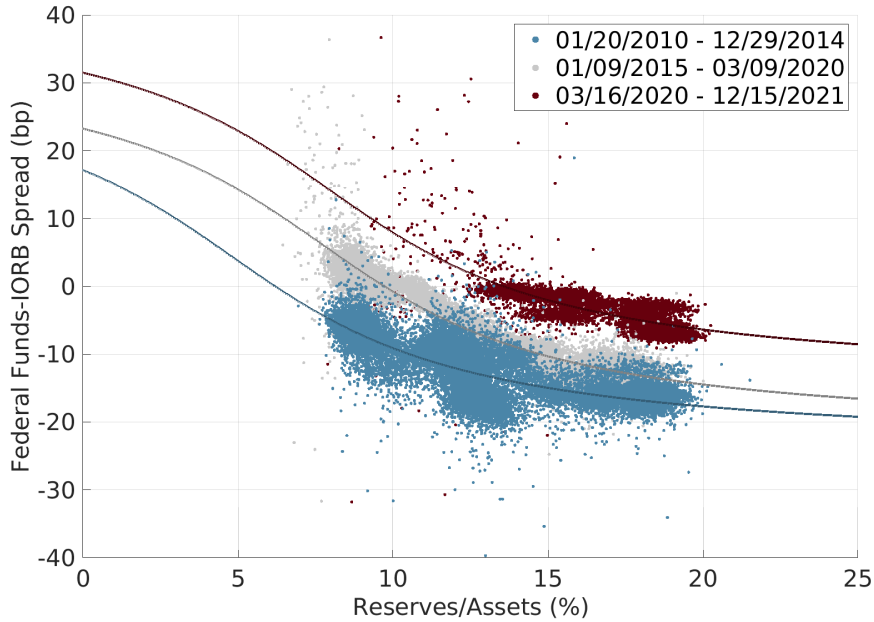


Figure IA.19: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data.** This figure shows the results of the nonlinear least-squares (NLLS) minimization in equation (IA.4). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of the time-varying VAR model in (A.8), with ten lags ($m = 10$ days). Forecasts are generated every five days; for each day, we draw $N = 100$ forecasts from the model-implied posterior joint distribution. Results are obtained constraining the fit to the right of the point of maximum absolute slope of the nonlinear demand function in (IA.3).

In terms of the time-invariant nonlinear part of the reserve demand function, our estimates show that the point of maximum slope (θ_1) occurs when reserves are around 5% of banks' assets; the point of maximum slope growth ($\theta_1 + \theta_2/\sqrt{3}$), instead, is around 8%. This estimate suggests that, in 2010-2014, the transition between ample and scarce reserves occurred around 8% of banks' assets. In 2015-2019, as a result of the 3-pp shift to the right q_2^* , this transition point seems to move to 11%. Finally, the normalization parameter θ_3 that captures the distance between the upper and lower asymptotes of the demand curve is 18 bp, which is close to the average DW-IORB spread in our sample divided by π (15 bp), confirming that our results are reasonable.

Looking back at the path of realized reserves over 2010-2021, these results suggest that, in both 2010 and 2019, with reserves between 8% and 10% of banks' assets, the federal funds market may have been operating around the transition between ample and scarce reserves;

in the second half of 2019, in particular, with reserves consistently below 9% of banks' assets and the threshold between ample and scarce around 11%, the market may have been operating inside the scarcity region.

	1/2010-12/2014	01/2015-03/2020	03/2020-12/2021
Horizontal shifts q^* (pp)	0	2.88	3.15
Vertical shifts p^* (bp)	-24.20	-22.31	-14.37

Table IA.6: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data - estimates of the shifts.** This table shows the estimates of the horizontal (q^*) and vertical (p^*) shifts in equation (IA.3) from the nonlinear least-squares (NLLS) minimization in equation (IA.4). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of our time-varying VAR forecasting model. The model includes ten lags ($m = 10$ days). Forecasts are generated every five days; for each day, we draw $N = 100$ forecasts from the model-implied posterior joint distribution. Results are obtained constraining the fit to the right of the point of maximum absolute slope of the nonlinear demand function in (IA.3).

Parameter	θ_1 (%)	θ_2 (%)	θ_3 (bp)
	4.78	5.69	18.25

Table IA.7: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data - estimates of the nonlinear time-invariant parameters.** This table shows the estimates of $\theta = (\theta_1, \theta_2, \theta_3)$ in equation (IA.3) from the nonlinear least-squares (NLLS) minimization in equation (IA.4). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of our time-varying VAR forecasting model. The model includes ten lags ($m = 10$ days). Forecasts are generated every five days; for each day, we draw $N = 100$ forecasts from the model-implied posterior joint distribution. Results are obtained constraining the fit to the right of the point of maximum absolute slope of the nonlinear demand function in (IA.3).

Internet Appendix G.4 Algorithm details

To solve the minimization problem in (IA.4), we use the `fmincon` function in MatLab. We choose the interior-point algorithm and set the maximum number of function evaluations to 10^9 , the maximum number of iterations to 10^{12} , the step tolerance to 10^{-18} , and the tolerance

on constraint violations to 10^{-12} . We analytically derive the gradient of the objective function in (IA.4) and include it in the minimization program.

To ensure that our results are reliable and economically meaningful, we set bounds on the variables of the minimization problem. We are agnostic about the origins and possible magnitude of the horizontal shift, q^* , in the reserve demand curve; for this reason, in all periods, we set its upper and lower bounds equal to two and minus two times the maximum level of normalized reserves in our sample, respectively. For the vertical shift p^* , we choose bounds based on the discussion of its economic interpretation in Internet Appendix G.2. The lower bound is the same for all periods and is equal to the minimum ONRRP-IORB spread in our sample; the rationale behind this choice is that the ONRRP rate is the safe outside option for FHLBs and MMFs, the main lenders to banks in the wholesale overnight funding market. The upper bound changes across periods and is equal to the average realized federal funds-IORB spread in the period; the reason is that $f(x; \theta)$ in equation (IA.3) is strictly positive everywhere, which implies that $p_t^* < p_t$ at all times by construction.

For the θ parameters governing the nonlinearity of the demand curve in (IA.3), we impose the following bounds. θ_1 represents the point of maximum absolute slope of the demand curve, where reserves are highly scarce. To bound θ_1 from below, we calculate the ratio between the aggregate reserve requirement and banks' total assets on each day, compute the minimum value in our sample, and then take 10% of that value. To bound θ_1 from above, we multiply the maximum of normalized reserves in our sample by two. We use the same bounds for θ_2 , which measures the distance between the point of maximum absolute slope and the point of maximum slope growth (i.e., the transition point between scarce and ample reserves) in the nonlinear function in (IA.3).

The upper asymptote of the nonlinear function (IA.3) is equal to $p_t^* + \pi\theta_3$. In the federal funds market, absent stigma, frictions, and caps on discount-window (DW) borrowing, the federal funds rate should be bounded from above by the DW rate. Therefore, given our bound on p^* , we set the upper bound on θ_3 to be equal to the maximum of the spread between the DW and ONRRP rates in our sample divided by π . The lower bound on θ_3 is simply set to zero.

To initialize the variables, we also build on our discussion of their economic interpretation in Internet Appendix G.2. We initialize θ_1 with the average level of reserves in 2009, which is the period of lowest reserve balances since the 2008 crisis; reserves in 2009 were most likely scarcer than in the rest of our sample. To initialize θ_2 , we exploit the experience of September 2019, which suggests that reserves may have transitioned from ample to scarce around that time (Afonso et al., 2021). Since the point of maximum slope growth in (IA.3)

is $\theta_1 + \theta_2/\sqrt{3}$, we initialize θ_2 with $\sqrt{3}(q_{2019} - \theta_1^{(0)})$, where q_{2019} is the average value of reserves in September 2019, and $\theta_1^{(0)}$ is the initialization of θ_1 . The initial value of θ_3 is set equal to the average spread between the DW and IORB rates in our sample divided by π . Data on the daily DW rate is publicly available from FRED (“DPCREDIT”).

Finally, we initialize q^* to zero in each subperiod, as we don’t have strong priors on its path over time; in each subperiod, we initialize p^* to one basis point below the minimum federal funds-IORB spread in that period, as p^* is strictly smaller than p by construction.

Parameters	Lower Bound	Upper Bound	Initial Value
θ_1	0.18 pp	38.83 pp	7.30 pp
θ_2	0.18 pp	38.83 pp	2.94 pp
θ_3	0.00 bp	25.5 bp	14.74 bp
q_t^*	-38.83 pp	38.83 pp	0.00 pp
p_t^*			
01/20/2010-12/29/2014	-25.00 bp	-12.52 bp	-20.00 bp
01/09/2015-03/09/2020	-25.00 bp	-6.33 bp	-15.00 bp
03/16/2020-12/15/2021	-25.00 bp	-3.31 bp	-8.36 bp

Table IA.8: **Bounds and initializations of the variables in the nonlinear least squares (NLS) minimization in equation (IA.4).** θ_1 , θ_2 , and θ_3 are the parameters defining the nonlinear time-invariant functional form of the reserve demand function in equation (IA.3); q^* and p^* are the horizontal and vertical shifts.