STAFF
REPORTS

NO. 1019
MAY 2022

REVISED
APRIL 2024

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

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

FEDERAL RESERVE BANK *of* NEW YORK

**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 April 2024
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 United States over 2010-21 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: reserves are abundant, and the demand curve is flat; 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. 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 (corresponding author), Williams: Federal Reserve Bank of New York (email: gabriele.laspada@ny.frb.org). Giannone: Amazon.com, Inc. The authors thank Marco Cipriani, Marco Del Negro, Darrell Duffie, Thomas Eisenbach, Huberto Ennis, Antoine Martin, Andrea Tambalotti, 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 Federal Reserve System Conference on Financial Institutions, Regulation, and Markets, and participants at numerous seminars for valuable feedback. They also thank Valerie Baldinger, Logan Casey, Peter Prastakos, Eric Qian, and Mihir Trivedi for excellent research assistance. Giannone's contribution was part of a continued collaboration based on work done prior to joining Amazon.

# 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 and Vandeweyer, 2021).

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 12 years of data, from 2010 to 2021.

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, transitioning

from steeply downward-sloping to flat as the supply of reserves increases. Second, the demand for reserves is not constant 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 nonlinear 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 nonlinear 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 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. Using daily data allows us to control for the FOMC's response to rate dislocations and for short-term disturbances to demand related to factors outside the Federal Reserve's control.

Our structural 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. Importantly, we obtain qualitatively similar results if we normalize reserves by bank deposits or GDP, confirming that the choice of the normalization factor does not 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 provide evidence of both vertical and horizontal low-frequency shifts in the demand for reserves over time. 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.

Finally, we show that our methodology can be used in real time to monitor the market for reserves and assess the relative ampleness of the supply of reserves. For example, if used in real time during the Federal Reserve's previous balance-sheet normalization, our methodology would have suggested 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.

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; 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., 2021; d'Avernas and Vandeweyer, 2021, 2023) and the implications of central bank balance-sheet expansions and contractions (Stein, 2012; Kashyap and Stein, 2012; Greenwood et al., 2016; Diamond et al., 2020; Acharya and Rajan, 2022; 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 identification strategy. Section 5 reports the results of our instrumental variable estimation. Section 6 reports the results of a post-processing fit of the demand curve. Section 7 concludes. The Appendix includes an extension of the model with banks' balance-sheet costs, a detailed description of the forecasting model, its out-of-sample validation, and robustness checks.

# 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 $\widetilde{r}_i$ 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 stochastic balance equal to $r_i' \equiv \widetilde{r}_i + f_i + z_i$, where $z_i$ is distributed according to a distribution function $G$ with support $(-\underline{z}, \overline{z})$. For simplicity, we assume that $G$ is absolutely continuous 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

[1] More 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).

[2] We 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.

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

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,

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

where $\hat{z}_i \equiv \bar{r}_i - \widetilde{r}_i - f_i$. 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 $f_i^*$:

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

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 $r_i = \widetilde{r}_i + f_i^*$ 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 $[\bar{r}_i - \overline{z}, \bar{r}_i + \underline{z}]$ and flat outside this interval, with $i^f(r_i) = i^{DW}$ for $r_i \leq \bar{r}_i - \overline{z}$ and $i^f(r_i) = i^{IORB}$ for $r_i \geq \bar{r}_i + \underline{z}$.

These properties of a bank's demand function carry over to the aggregate (inverse) demand for reserves of the banking system. We can derive the aggregate demand 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 summing reserves across banks to obtain the aggregate demand $r \equiv \sum_{i=1}^N r_i = \sum_{i=1}^N \bar{r}_i - NG^{-1}\left(\frac{i^f - i^{IORB}}{i^{DW} - i^{IORB}}\right) \equiv \bar{r} - NG^{-1}\left(\frac{i^f - i^{IORB}}{i^{DW} - i^{IORB}}\right)$.[5] Inverting

[4] It 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.

[5] If $G$ has bounded support, aggregate demand is $r \leq \sum_{i=1}^N (\bar{r}_i - \overline{z}) = \bar{r} - N\overline{z}$ 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 $r \geq \sum_{i=1}^N \bar{r}_i + \overline{z} = \bar{r} + N\underline{z}$ for $i^f = i^{IORB}$.

this expression back, we obtain the aggregate (inverse) demand curve for reserves:

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

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 assuming no frictions. In practice, however, there are frictions that limit banks' ability to arbitrage and therefore affect banks' demand for reserves. Banks' costs based on balance-sheet size are among the most important ones. These costs penalize borrowing and balance-sheet expansions, especially to finance safe assets such as reserves, limiting banks' ability to take arbitrage positions (Anderson et al., 2020; Copeland et al., 2021). Notable examples are the revised Federal Deposit Insurance Corporation (FDIC) assessment fee and the Basel III leverage-ratio regulation introduced after the 2008 financial crisis. As a result of these costs, when borrowing to run 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 run the DW arbitrage, banks demand a higher rate than the rate at which they borrow from the central bank.[7] In Appendix A, we present a model extension that

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

[7] Another 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).

incorporates banks' balance-sheet costs; see also Afonso et al. (2019) and Kim et al. (2020).

Another important friction in the federal funds market is segmentation. Federal Home Loan Banks (FHLBs)—a type of government-sponsored enterprise and the main lenders in this market—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.[8] As a result, the lower bound of the reserve demand curve lies between the IORB and ONRRP rates, with the distance from these bounds depending on the bargaining power of FHLBs relative to borrowing banks. Since our goal is to estimate the slope of the demand curve, and this friction does not affect the shape of the curve but only the vertical location of its lower bound, our stylized model provides a valid theoretical framework to guide our empirical analysis. See Afonso et al. (2019) and Lagos and Navarro (2023) for models that incorporate this market segmentation.

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 in 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.

Knowing the shape of the reserve demand curve is key for monetary policy because the FOMC uses the federal funds rate to communicate the policy stance. As many other central banks after 2008, the Federal Reserve operates a floor system in which the supply of reserves is sufficiently large that typical reserve fluctuations do not move prices significantly, and control over short-term rates is implemented through changes in the IORB and other administered rates (FOMC, 2019). To successfully implement such a system, it is important to identify the transition between abundant and ample reserves, that is, the demand satiation point. A supply of reserves below such point may lead to severe price dislocations (Correa et al., 2020; Afonso et al., 2021; Copeland et al., 2021; d'Avernas and Vandeweyer, 2021).

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 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

[8]The Federal Reserve adjusts the ONRRP and IORB rates in the implementation of monetary policy. See Cipriani and La Spada (2022) for details.

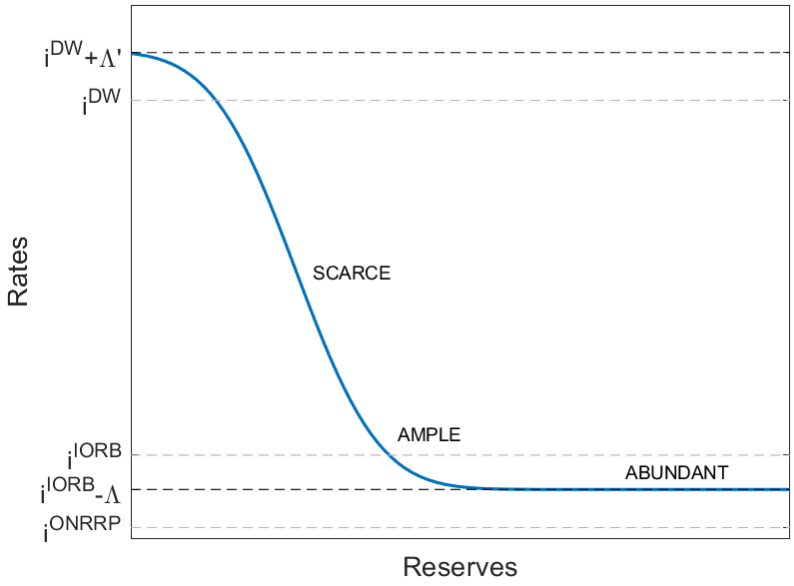


Figure 1: **Bank demand for reserves in the presence of market frictions.** Example of reserve demand curve with banks' balance-sheet costs, market segmentation, and discount-window stigma. $i^{DW}$, $i^{IORB}$, and $i^{ONRRP}$ represent the DW, IORB, and ONRRP rates. $\Lambda$ captures the effects of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

and flattens as reserves increase. (4) Variation in banks' balance-sheet costs or in FHLBs' bargaining power move the vertical location (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 Endogeneity

To understand the sources of endogeneity in the demand for reserves, it is important to understand how aggregate reserves can change. Reserves in the banking system change for two reasons: either because the Federal Reserve buys assets from or sells them to banks, which changes the size of its balance sheet, or because funds are transferred between reserves and non-reserve accounts at the Federal Reserve, which changes the composition of its liabilities. Some institutions have Federal Reserve accounts and transact with banks; when these transfers occur, reserves change outside the Federal Reserve's control.[9] Both types of

[9]In contrast, federal funds transactions between banks just redistribute reserves within the system.

reserve fluctuation are associated with endogeneity problems.

The first type of endogeneity is due to the Federal Reserve's actions. The Federal Reserve on occasion responds to unusual volatility in the federal funds rate by adjusting the supply of reserves to keep the rate within its target range, as occurred in September 2019 and March 2020.[10] These responses are quick and put in place within a matter of days. In September 2019, for example, the federal funds rate spiked upwards on the $16^{th}$ and breached the top of its target range on the $17^{th}$; starting on the $18^{th}$, the Federal Reserve expanded the supply of reserves, and rates returned to their prior levels (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, their demand for short-term funding increases as they finance their purchases with overnight repos. The temporary increase in repo rates can increase the demand for reserves because overnight federal funds and repos are close substitutes (Schulhofer-Wohl and Clouse, 2018).[11]

Martin et al. (2019) provide another example of how Treasury issuance, and therefore activity in the TGA, can affect the federal funds market. The issuance of Treasuries 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 increases the demand for reserves, pushing the federal funds rate up.

Another important example of non-reserve liability correlated with banks' demand for reserves is the ONRRP. Through this facility, both FHLBs and MMFs—the main lenders

[10]Before 2008, when reserves were very scarce, this type of endogeneity was more severe because the Federal 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.

[11]Another confounding factor related to the TGA are tax payments. When money market fund 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, to meet redemptions, money funds reduce their overnight repo lending, which can lead to a temporary rise in repo rate, increasing the demand for reserves and the federal funds rate. Both Treasury settlements and tax payments played important roles in the money-market turmoil of September 2019 (Afonso et al., 2021).

in the federal funds and repo markets, respectively—can invest at the Federal Reserve via overnight repos at a fixed rate below the IORB. One way variations in the ONRRP balance correlate with the demand for reserves is through the "window dressing" of European banks around month-ends. In Europe, the Basel III leverage ratio is calculated using only month-end data, which gives European banks an incentive to temporarily reduce their overnight borrowing around those dates, both in the federal funds and in other money markets (Banegas and Tase, 2020). 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 the ONRRP to compensate the reduction in funding demand. When MMFs invest in the ONRRP, they instruct their custodian banks to make the transfer; the result is a decrease in bank reserves and an equal increase in the ONRRP.

More generally, changing conditions in the repo and Treasury markets affect not only banks' demand for reserves by changing the price of substitute forms of overnight funding, but also the supply of reserves by changing MMFs' incentives to invest in the ONRRP. Overnight Treasury repos have become particularly relevant in both banks' funding demand and MMFs' lending supply since the 2014 SEC reform of the MMF industry (Anderson et al., 2020; Cipriani and La Spada, 2021).[12]

The endogeneity due to non-reserve Federal Reserve liabilities has become more important since 2008. First, the repo and Treasury markets have grown substantially. In the tri-party repo market, for example, lending collateralized by Treasuries and agency debt increased from $1.3 trillion in May 2010 to $3.3 trillion in December 2021.[13] Second, the variability of non-reserve liabilities has also increased significantly (Afonso et al., 2020). Figure 2 shows that, excluding currency in circulation, non-reserve liabilities amounted to 32% of reserves in January 2010 and to 63% in December 2021, with a peak of 89% in July 2020. The TGA increased from $90 billion at the end of 2010 to $284 billion at the end of 2021, with a peak of $1.8 trillion in July 2020. Since its inception, the ONRRP has fluctuated between less than a billion and $1.7 trillion, reached in December 2021.[14]

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.

[12]The reform led investors to move more than $1 trillion from prime MMFs, which can lend to banks both unsecured and secured, to government MMFs, which can only lend to banks via Treasury and agency repos.

[13]See `https://www.newyorkfed.org/data-and-statistics/data-visualization/tri-party-repo`.

[14]Afonso et al. (2022) discuss the drivers of the recent surge in ONRRP investment.

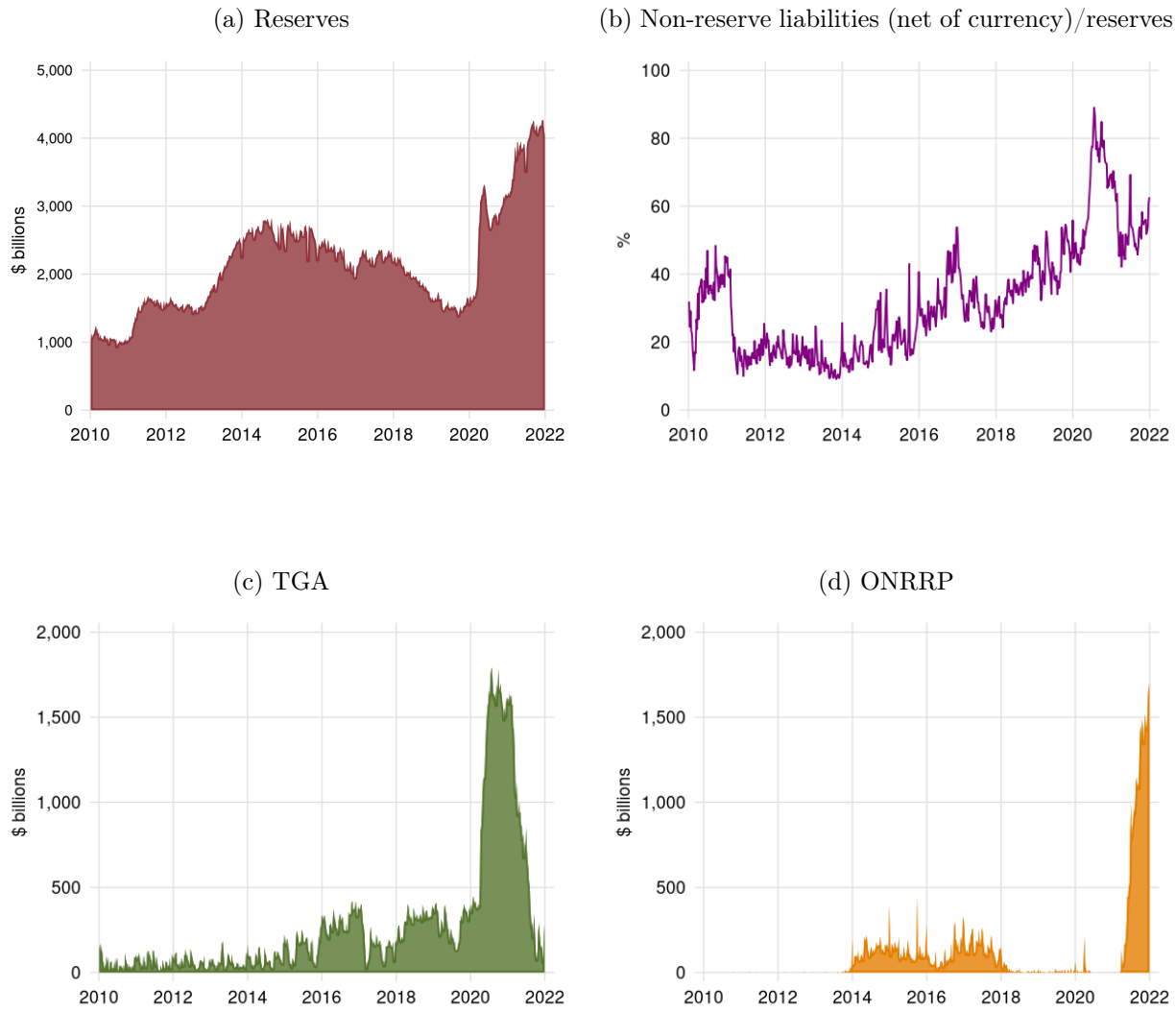


Figure 2: **Federal Reserve Liabilities from January 2010 to December 2021.** Reserves (panel (a)), ratio of non-reserve liabilities (excluding currency in circulation) to reserves (panel (b)), Treasury General Account (TGA) (panel (c)), and Overnight Reverse Repurchase Agreement (ONRRP) facility (panel (d)). Data are weekly from Federal Reserve Statistical Release H.4.1.

# 3 Data

## 3.1 Data description

We use daily data from January 1, 2009 to December 29, 2021. 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 month-ends. 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 federal funds rates. To calculate reserves, we use daily confidential data on aggregate balances held by depository institutions, provided by the Federal Reserve Bank of New York.[16] To take into account the growth of the banking system over time, we normalize reserves by commercial banks' assets. Weekly data on banks' assets are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). We linearly interpolate these weekly data to obtain a daily series.

For the federal funds rate, we use the daily volume-weighted average based on transactions data, collected by the Federal Reserve Bank of New York. 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 by the Fixed Income Clearing Corporation (FICC), which is publicly available on the website of the Depository Trust & Clearing Corporation (DTCC).[17] 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").

[15] We 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`.

[16] A weekly version is available in the H.4.1 report ("Reserve balances with Federal Reserve Banks").

[17] `https://www.dtcc.com/charts/dtcc-gcf-repo-index`.

### 3.2 Preliminary evidence

Panel (a) of Figure 3 shows the time evolution of aggregate reserves normalized by banks' total assets. Reserves went through a full expansion-contraction cycle from 2010 to late 2019 and expanded again in early 2020 until late 2021, ranging from 8% (2010 and 2019) to 19% (2014 and 2021) of banks' assets. These movements reflect the Federal Reserve balance-sheet expansions in response to the 2008 and 2020 crises, as well as the interim normalization period (2015-2019). 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, at the onset of the Covid pandemic.

Figure 3 simply shows equilibrium realizations over time and cannot be interpreted causally. To identify the slope of the reserve demand curve, we need our structural time-varying methodology.

## 4 Empirical Implementation

### 4.1 Empirical model and estimation

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

$$p_t = p_t^* + f\left(q_t - q_t^*; \theta_t\right), \tag{4}$$

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 $\theta$ 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-

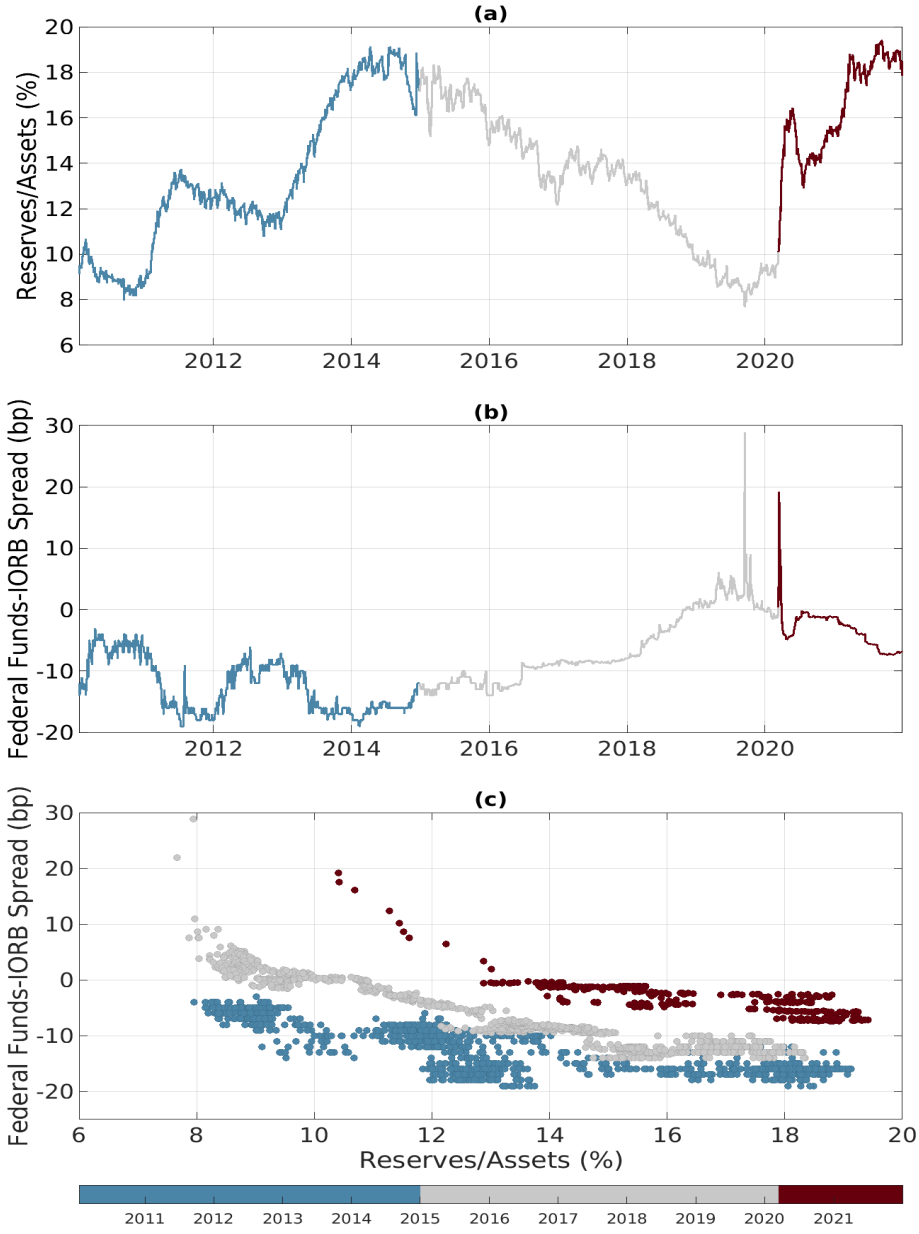


Figure 3: **Reserves, the federal funds rate, and the reserve demand curve.** Panel (a) plots aggregate reserves relative to commercial banks' assets from January 1, 2010 to December 29, 2021. Panel (b) shows the spread between the volume-weighted average federal funds rate and the IORB rate (in basis points). Panel (c) plots the relationship between the spread and normalized reserves. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

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 bargaining power of FHLBs (see Section 2.1). Variation in horizontal location of the curve $q_t^*$ comes from structural changes in 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. All parameters, including the shock variance $\sigma$, can vary at daily frequency. The time-varying slope $\beta_t$ 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 a straight line; as time passes, these lines move and trace the demand curve. This approach enables us to capture the nonlinear nature of the demand curve, without specifying a particular 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 $\beta$ from (high-frequency) variation in $v$. The assumption is plausible for two reasons. First, it took banks months to adjust to the changes in regulation, supervision, and market functioning that occurred after 2008. Second, as we explain below, our daily movements along the curve are small, so that the locally linear approximation works well, and reflect exogenous variation in the expansions and contractions of the Federal Reserve balance sheet, which took place over

many years.

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 the work of Hamilton (1997), who measures the slope of the reserve demand curve in 1989-1991 using forecast errors from a time-invariant model as instrument.

The idea is to use variation in reserves that is residual to the Federal Reserve policy response to dislocations in the federal funds rate and uncorrelated with the transitory confounding factors due to non-reserve Federal Reserve accounts.

Our forecasting model is the following time-varying vector autoregressive (VAR) model with stochastic volatility based on Primiceri (2005) and Del Negro and Primiceri (2015):

$$\begin{aligned} q_t &= c_{q,t} + b_{q,q,1,t} q_{t-1} + b_{q,p,1,t} p_{t-1} + ... + b_{q,q,m,t} q_{t-m} + b_{q,p,m,t} p_{t-m} + u_{q,t}, \\ p_t &= c_{p,t} + b_{p,q,1,t} q_{t-1} + b_{p,p,1,t} p_{t-1} + ... + b_{p,q,m,t} q_{t-m} + b_{p,p,m,t} p_{t-m} + u_{s,t}, \end{aligned} \tag{6}$$

where $q$ and $p$ are the quantity and price of reserves—reserves over assets and federal funds rate minus IORB rate—the $c$'s and $b$'s are the time-varying coefficients, and the $u$'s are the forecast errors. These errors are serially uncorrelated and jointly normally distributed with mean zero and time-varying covariance matrix $\Omega_t$, i.e., $(u_{q,t}, u_{p,t})' \sim \mathcal{N}(0, \Omega_t)$.

We estimate model (6) 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 model (6) is that its parameters evolve more slowly than the daily errors. Like Primiceri (2005), we assume that the parameters follow slow-moving random walks, whose innovations are uncorrelated with the $u$ errors at all leads and lags. Given the data, we estimate the joint posterior distribution of the $c$'s, $b$'s, and $\Omega$ on each day. Our IV estimation then maps these parameters into the slope of the demand curve.

Using the forecast error for reserves $h$ days ago ($u_{q,t-h}$) as an instrument for reserves today ($q_t$) in equation (5), we can write the IV estimate of $\beta_t$ as

$$\beta_t^{IV} = \frac{\text{cov}(p_t, u_{q,t-h})}{\text{cov}(q_t, u_{q,t-h})}. \tag{7}$$

Our estimation of the forecasting model (6) gives us simultaneously the forecast errors $u$ and their time-varying covariances with the observable variables $q$ and $p$. These covariances

are functions of the time-varying model parameters and can be obtained by drawing from their posterior distribution. In Appendix B.1, we describe in detail how to conduct inference on $\beta_t^{IV}$ and explain how the ratio of the covariances in equation (7) can be interpreted as the ratio of the $h$-day-ahead impulse responses of rates and reserves to the forecast error of reserves.[18]

Finally, inference on $\beta_t^{IV}$ depends on the choice of the forecast horizon $h$. There is a clear trade-off between instrument exogeneity and estimate precision. The longer is the horizon, the more plausible is the exogeneity assumption. A longer horizon, however, implies larger estimate uncertainty. As we discuss in the next section, we use $h = 5$ (i.e., one week) because, based on the institutional details of the market for reserves, this choice should satisfy the exogeneity requirement, while keeping our estimates sufficiently precise.

## 4.2 Exogeneity and relevance of our instrument

The instrument is exogenous if the confounding factors, related to either the Federal Reserve's response to rate dislocations or factors outside the Federal Reserve's control, have transitory effects disappearing within a week. The exogeneity assumption underlying our identification strategy is that the forecast error for reserves at time $t$ from model (6) is uncorrelated with the shock at time $t + 5$ in the demand equation (5). That is, an error in the forecast of reserves only affects banks' (future) demand for reserves through its effect on the (future) level of reserves.

The exogeneity of our instrument is plausible for two reasons. First, the FOMC's reserve supply does not react on the same day to fluctuations in the federal funds rate; it responds quickly but with a delay of at least a day. By conditioning on past daily data up to ten lags and allowing for time-varying coefficients, our forecasting model can control for the FOMC's response function to price dislocations; as a result, its errors should be residualized with respect to this confounding factor. For this reason, it is important to use daily data to extract exogenous variation in the supply of reserves. Variation in reserves based on average data at lower frequencies, say weekly or monthly, would reflect the FOMC's response to demand shocks.

Second, the effect of factors outside the Federal Reserve's control that are simultaneously correlated with both reserves and banks' demand for them are likely to last less than five

[18]The ratio of impulse response functions as an IV estimator was 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. Impulse response functions (IRFs) relative to forecast errors are also known as generalized IRFs (Koop et al., 1996).

business days. The settlements of Treasury auctions and tax payments are good examples. Although these events affect aggregate reserves through changes in the Treasury's account at the Federal Reserve and are correlated with banks' demand for short-term borrowing, they are transitory in nature. Therefore, even if today's forecast error may be correlated with today's demand shock due to the activity of these non-reserve Federal Reserve accounts, the error from five days ago should not be.

Finally, to further strengthen our identification, our robustness checks also include daily repo and Treasury bill rates in the VAR forecasting model. In this way, our IV estimation directly controls for possible spillovers from the repo and Treasury markets to both the federal funds rate and aggregate reserves. The estimation of these trivariate VAR models is similar to that of the baseline bivariate model; details can be found in Appendix B.1.

The relevance of our instrument stems from the persistence of the reserve path in our sample and the forecasting accuracy of our model. From 2010 to 2019, the Federal Reserve gradually expanded and contracted the supply of reserves and then expanded it again in March 2020 in response to the Covid-19 crisis. These trends, which last for months and around which higher-frequency events can occur, are well captured by the autoregressive nature of our forecasting model; as we formally show in Appendix B.4, in fact, our model displays reasonable out-of-sample predictive accuracy. As a result, the time-varying covariance of reserves five days ahead with their forecast error today (i.e., the denominator in the IV estimate (7)) is significant throughout our sample (see panel (b) of Figure 6). This covariance measures the relevance of our instrument and can be interpreted as the result of the first-stage regression in the two-stage least-squares (2SLS) estimation.

An important advantage of our approach relative to ordinary 2SLS is that we do not need to test for instrument strength because our inference is automatically robust to weak instruments. The reason is that the Bayesian posterior distribution of $\beta_t^{IV}$ already reflects the uncertainty in both the numerator (i.e., the reduced-form coefficient) and denominator (i.e., the first-stage coefficient). In this way, our inference directly accounts for the possible non-normality of the IV estimate (7), which could occur if the denominator is close to zero.

Our IV estimation is also robust to autocorrelation and heteroscedasticity of the demand shocks in equation (5). The reason is that the elasticity $\beta_t$ is derived as a function of time-varying covariances estimated from a VAR model with ten lags and stochastic volatility.

### 4.3 Predictive accuracy of our model

Our forecasting model (6) is very flexible and general because it allows all the parameters, including the variances and covariances of the residuals, to change over time. Thanks to this feature, the model can account for complex dynamic interactions between the federal funds rate and reserves. In many contexts, high model complexity comes with the curse of dimensionality: the model tends to overfit in-sample and perform poorly out-of-sample. Several papers, however, have shown that, when used for macroeconomic forecasting with quarterly data, the class of models encompassing model (6) does not suffer from this pathology and produces accurate forecasts even in real time (D'Agostino et al., 2013). The reason is that the priors specified by Primiceri (2005) are very conservative in the amount of time variation.

This is true also in our context with daily financial data. Figure 4 reports in-sample and out-of-sample joint forecasts of the federal funds rate and reserves five days ahead. Model (6) is not affected by the curse of dimensionality: the in-sample and the out-of-sample predictions are similar and close to the realized data. In Appendix B.4, we formally evaluate the out-of-sample accuracy of the model's predictions in real time under various metrics. We show that the model provides reliable and stable inference. Predictive accuracy is a key advantage of our framework as it enables us to monitor in real-time the relative ampleness of reserves in the banking system, as we discuss in Section 5.3.

## 5 Empirical Results

### 5.1 Suggestive evidence from the forecasting model

Figure 4 provides qualitative suggestive evidence of the two main results of the paper: (i) the reserve demand curve is nonlinear with a satiation point, and (ii) it has shifted outward over the last ten years, both horizontally and, especially, vertically.

To reflect the different cycles of expansion and contraction of the Federal Reserve balance sheet, we split our sample in Figure 4 in three periods: from 2010 to 2014 (expansion), from 2015 to mid-March 2020 (contraction), and from mid-March 2020 to December 2021 (expansion). Consistent with economic theory, and model (3) in Section 2.1, the relationship between the federal funds rate and reserves is nonlinear in all three periods: it is negative for levels of reserves that are sufficiently low, say below 10-11% of banks' assets, whereas it is flat when reserves are sufficiently large, say above 14-15% of banks' assets.

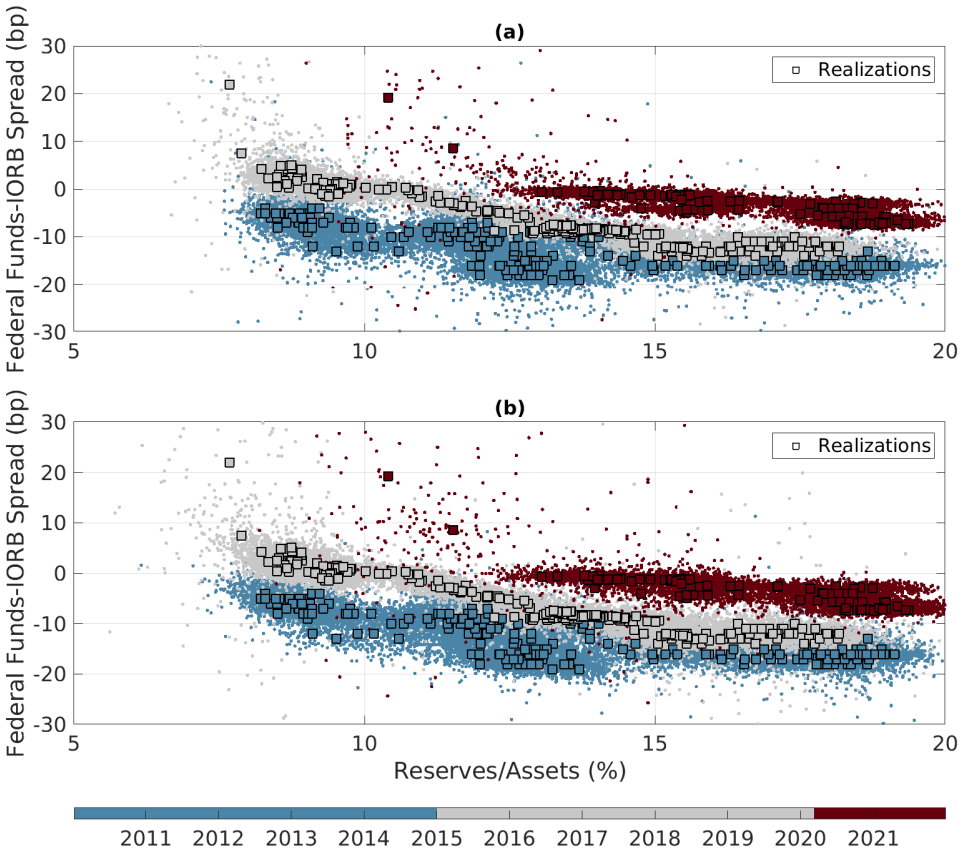


Figure 4: **In-sample (a) and out-of-sample (b) joint forecasts of the federal funds rate and reserves five days ahead.** The forecasts are generated using the bivariate time-varying model (6), drawing 100 times every five business days from the joint posterior distribution of reserves and rates for each day. The black squares represent the realized data on the day for which forecasts are generated (i.e., $t + 5$). Reserves are measured as a ratio to commercial banks' assets. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

Has the location of the reserve curve shifted over time? The level of reserves at which the relationship between prices and quantities transitions from being flat to negative seems to slightly change over time, even after adjusting reserves for the growth of the banking system. During 2010-2014, the slope is negative for reserve levels below 11-12% of banks' assets, which roughly correspond to 2010 and 2011; during 2015-2019, a negative slope of similar magnitude emerges for reserve levels below 13-14% of banks' assets, which correspond to 2018 and 2019. The seemingly higher threshold between the sloped and flat regions in the second half of the sample suggests a modest horizontal shift to the right.

More importantly, over 2010-2021, the relationship between the federal funds rate and aggregate reserves seems to mainly move vertically. As shown in Figure 4, the 2015-2020Q1 curve is above the 2010-2014 curve not only in the sloped region but also in the flat one, which would not happen if the curve had only moved horizontally. In particular, the 2015-2017 rates tend to be consistently above the 2012-2014 ones at every level of reserves, although the curve is flat in both periods.

A second and even more relevant vertical shift is visible for the last period. From March 2020 to December 2021, federal funds rates and their forecasts have been consistently above those from previous years, even by more than 10 bp, at every level of reserves; this is true also in the flat region of the curve, which represents most of this time period.

The next section presents our time-varying IV estimates of the slope of the reserve demand curve, which are consistent with a nonlinear demand function with a satiation level; in Section 6, we quantify the horizontal and vertical shifts through a post-processing methodology and discuss their possible economic interpretations.

## 5.2 Demand satiation and the slope of the reserve demand curve

### 5.2.1 OLS estimation

We now turn to estimating the time-varying slope of the reserve demand curve, i.e., the elasticity $\beta_t$ 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 forecasts from model (6) as pseudodata. Every five days, we draw $N = 2,500$ forecasts from the model-implied five-day-ahead joint distribution of spreads and reserves and run a pooled regression over the past year (244 days). Figure 5 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.

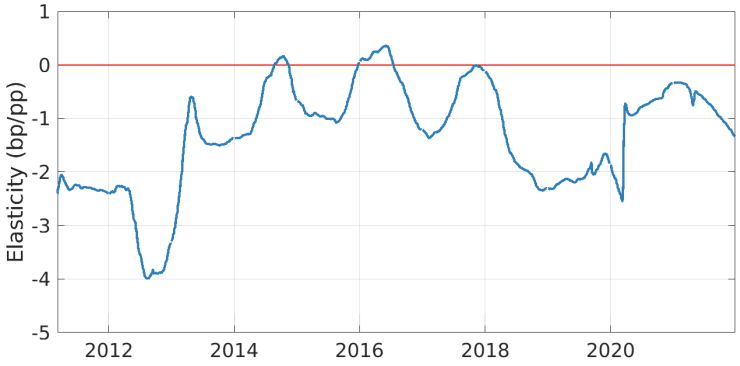


Figure 5: **OLS estimate of the elasticity of the federal funds rate to reserves from model forecast.** The slope of the reserve demand curve is estimated by running OLS regressions on rolling windows (244 business days) of in-sample forecasts of the spread between the federal funds rate and the IORB rate against in-sample forecasts of reserves normalized by banks' assets. The forecasts are generated using the bivariate time-varying model (6), drawing 100 times every five business days from the joint posterior distribution of the reserves and rates for each day.

### 5.2.2 IV estimation

The slopes in Figure 5, however, cannot be interpreted causally because those forecasts do not control for endogeneity. To address identification issues, we use the IV approach described in Section 4; Figure 6 presents the results. Panels (a) and (b) show the time-varying posterior medians of the numerator and denominator of our IV estimate in equation (7), together with their 95% and 68% confidence bands; panel (c) shows the same information for the IV estimate itself.

The numerator in equation (7), depicted in panel (a), is the time-varying covariance of rates 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. Figure 6 shows that, over time, the reduced-form coefficient and IV estimate closely move together, in terms of both sign and statistical significance, which reassures us of the validity of our inference.

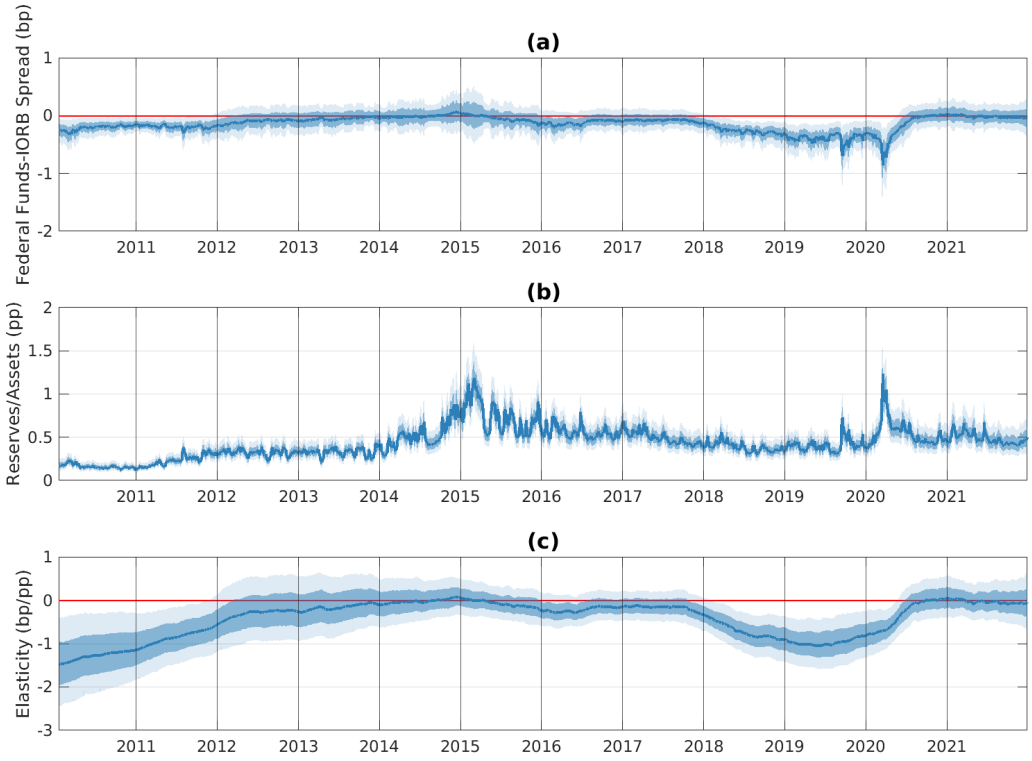


Figure 6: **IV estimate of the elasticity of the federal funds rate to reserves.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample from model (6) 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 commercial banks' assets. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

The denominator in equation (7) is the time-varying covariance of reserves and their past forecast errors, which can be interpreted as the strength of our instrument. Panel (b) of Figure 6 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 5: the time path of the IV estimate is similar to the time-varying slope from the rolling OLS regression on the model's forecasts. 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%, it 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. 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 of the slope of the demand function, $\beta_t^{IV}$, in that year. In 2010, a one-percentage-point drop in the ratio of reserves to banks' 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 rates' daily changes.

| 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (a) Bi-variate Model | | | | | | | | | | | |
| -1.27 | -0.84 | -0.31 | -0.15 | -0.01 | -0.07 | -0.21 | -0.16 | -0.68 | -0.96 | -0.28 | -0.03 |
| (-2.17,-0.31) | (-1.57,-0.11) | (-1.01,0.51) | (-0.81,0.58) | (-0.5,0.5) | (-0.47,0.36) | (-0.57,0.17) | (-0.53,0.2) | (-1.25,-0.13) | (-1.52,-0.36) | (-1.07,0.39) | (-0.49,0.47) |
| (b) Tri-variate Model with Repo Rates | | | | | | | | | | | |
| -1.92 | -1.3 | -0.84 | -0.64 | -0.14 | -0.13 | -0.22 | -0.17 | -0.62 | -0.98 | -0.3 | -0.12 |
| (-2.87,-0.98) | (-2.21,-0.41) | (-1.65,0.05) | (-1.52,0.18) | (-0.72,0.41) | (-0.56,0.34) | (-0.65,0.21) | (-0.6,0.27) | (-1.17,-0.09) | (-1.63,-0.36) | (-1.14,0.39) | (-0.62,0.41) |
| (c) Tri-variate Model with Treasury Yields | | | | | | | | | | | |
| -1.38 | -0.94 | -0.38 | -0.27 | -0.07 | -0.07 | -0.24 | -0.13 | -0.74 | -1.03 | -0.24 | -0.08 |
| (-2.35,-0.37) | (-1.76,-0.07) | (-1.24,0.58) | (-1.04,0.59) | (-0.64,0.52) | (-0.54,0.42) | (-0.67,0.2) | (-0.56,0.32) | (-1.37,-0.09) | (-1.62,-0.39) | (-1.1,0.49) | (-0.62,0.48) |



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 (7). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (6); in panel (b), they are estimated from an augmented trivariate version of model (6) that also includes daily repo rates; in panel (c), they are estimated in-sample from an augmented trivariate version of model (6) 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 commercial banks' assets. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York; daily data on repo rates are for overnight Treasury repos and available from the Depository Trust & Clearing Corporation (DTCC) at `https://www.dtcc.com/charts/dtcc-gcf-repo-index`. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances and daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER", "IORB", and "DGS1" respectively).

Taken together, the results in panel (c) of Figure 6 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 by 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 nominal reserves; see 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 time-varying 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 over time.

Still, for the interpretation of our findings, an important question remains: where does the low-frequency time 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 time variation in $\beta_t^{IV}$ 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 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.

The results of this section are important for monetary policy implementation because they inform policy makers 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 5.3, 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.

The results of this section, however, do not tell us: (i) whether there have been vertical structural shifts over time, as they do not affect the slope of the demand curve; and (ii) for what level of reserves the curve transitions from the ample to the scarce region, where the curve becomes increasingly steeper and reaches its maximum (negative) slope. In Section 6, we address both questions using a non-structural, post-processing approach.

#### 5.2.3 Robustness: Controlling for repo rates and Treasury yields

Model (6) may be misspecified as other factors could affect the relationship between the federal funds rate and aggregate reserves, such as the repo and Treasury markets. To explicitly control for the effect of repo-market conditions, we augment the forecasting model (6) by including the spread between the daily repo and IORB rates. We then use the reserve forecast errors generated by this trivariate time-varying VAR as the instrument for reserves in our IV estimation; as in our baseline specification, we use forecast errors from five days before. Figure 7 shows the results of the trivariate model, which are consistent with those of the bivariate one. We find that the reserve demand curve displays a negative slope in 2010-2011 and in 2018-April 2020, whereas its slope is statistically insignificant during the interim period between 2012 and 2017 and after April of 2020.[19]

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 consistent with 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 federal funds-IORB spread by 2 bp in 2010 and 1 bp in 2019; during 2012-2017 and 2020-2021, the slope is statistically insignificant.

To explicitly control for the possible confounding effect of Treasury-market conditions, we proceed in a similar fashion. We augment our forecasting model (6) with the spread between daily market yields on one-year Treasuries and the IORB rate. We then use the reserve forecast error from five days before as the instrument for reserves in our IV estimation.

[19]The only exception is the 2012Q4-2013Q2 period. A slight but steady decline in reserves begins in the second quarter of 2011, just after the second round of large-scale asset purchases (LSAPs), and lasts until the fourth quarter of 2012, when the third round of LSAPs starts. As a result, the slope becomes slightly negative in October 2012, but it returns to be indistinguishable from zero in June 2013, during the persistent balance sheet expansion of 2013.

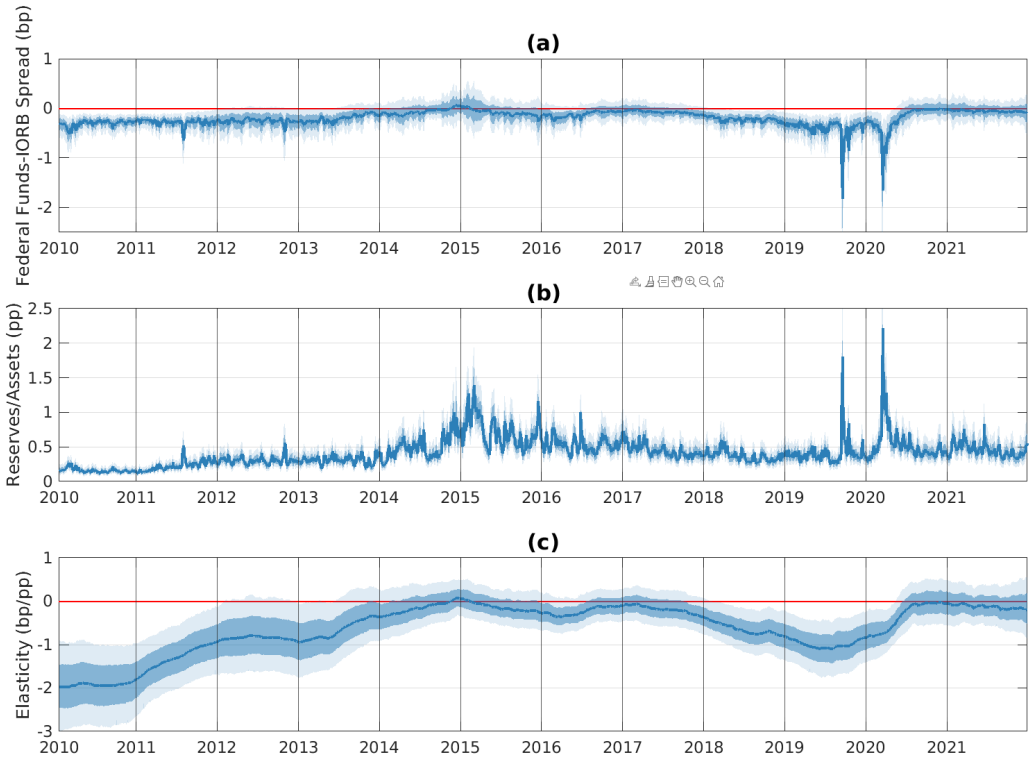


Figure 7: **IV estimate of the elasticity of the federal funds rate to reserves controlling for repo rates.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (6) 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 commercial banks' assets. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York; daily data on repo rates are for overnight Treasury repos and available from the Depository Trust & Clearing Corporation (DTCC) at `https://www.dtcc.com/charts/dtcc-gcf-repo-index`. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

Figure 8 shows the results for this specification. Consistent with the results of the bivariate model in Figure 6 and the trivariate model with repo rates in Figure 7, 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.

Quantitatively, the effect of a movement along the demand curve when controlling for Treasury yields is also very close to that obtained from the baseline specification. Panel (c) of Table 1 reports our estimates of the yearly elasticities: in 2010, a one-percentage-point decrease in normalized reserves leads to a median increase in the federal funds to IORB spread of 1.3 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.

## 5.3 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) elasticities from the baseline IV estimation using the bi-variate VAR model (6) to construct the reserve instrument. The IS estimate (blue) is the same as the one in Figure 6: 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. The OOS elasticity is what the econometrician would measure if they re-estimated the model every day, expanding the sample by one day; comparing in-sample and out-of-sample results tells us 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 almost 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

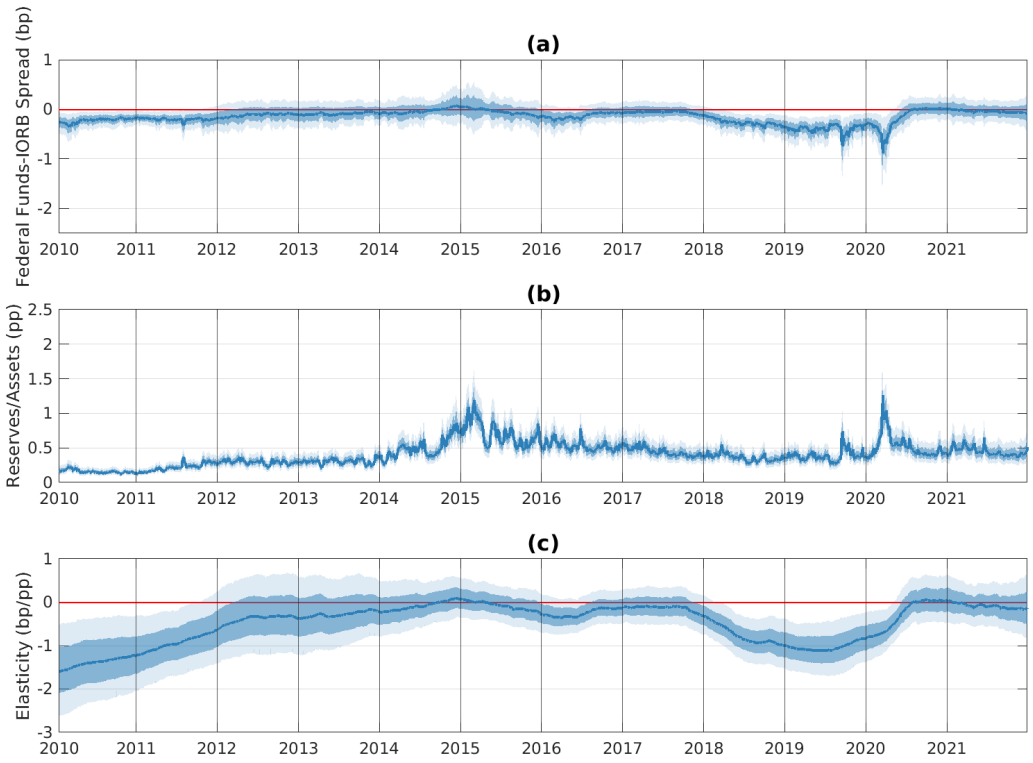


Figure 8: **IV estimate of the elasticity of the federal funds rate to reserves controlling for Treasury yields.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (6) 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 commercial banks' assets. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances and daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER", "IORB", and "DGS1" respectively).

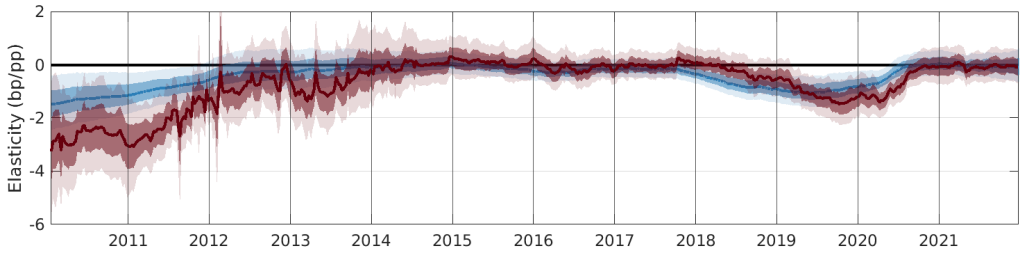


Figure 9: **IS and OOS IV estimates of the elasticity of the federal funds rate to reserves.** The blue solid line represents the posterior median IS estimate; the red solid line shows the posterior median OOS estimate. The dark and light shaded areas correspond to the 68% and 95% confidence bands. Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Weekly data on the total assets of U.S. commercial banks and U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

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 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. In particular, the OOS elasticity dips further down in late February, three weeks before the federal funds-IORB spread spikes to 19 bp on March 16 due to the money-market turmoil caused by the Covid crisis.[20] 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 demand or supply shocks could lead to significant price dislocations. In particular, for both the September 2019 and March 2020 stress episodes, we show that our time-varying elasticity estimates would have provided an early-warning signal several months in advance.

[20] In contrast, the IS elasticity, although significantly negative, remains roughly flat—which is not surprising as it is a smoothed estimate.

# 6 Implications for 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.

## 6.1 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).[21] As a result, absent vertical shifts, an increase in the spread would im-

[21] In 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).

ply 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 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:

$$p_t = p_t^* + f(q_t - q_t^*; \theta) \qquad \text{with} \quad f(x;\theta) = \left(\arctan\left(\frac{\theta_1 - x}{\theta_2}\right) + \frac{\pi}{2}\right)\theta_3, \tag{8}$$

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: $\theta_1$ is a location parameter, $\theta_2$ is a scale parameter, and $\theta_3$ 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 $x \to \infty$, as predicted by the theory (see equation (3)). Consistent with the evidence in Figure 4, we consider three periods, corresponding to different locations of the curve: 2010-2014, 2015-3/09/2020, and 3/16/2020-12/29/2021. 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 $\theta$ is constant across periods.

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:

$$\sum_{k=1}^{3} \sum_{t \in T_k} \sum_{i=1}^{N} \left[p_{it} - p_k^* - f(q_{it} - q_k^*; \theta)\right]^2 \tag{9}$$

where $T_1$, $T_2$, and $T_3$ represent 2010-2014, 2015-3/09/2020, and 3/16/2020-12/29/2021; $i = 1, \ldots, N$ 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 (6).[22] 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 (9).

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

## 6.2 Parameter interpretation and constraints

Our estimation method is silent regarding the origins of the vertical and horizontal shifts in (8). 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, $\theta_1$ 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 $\theta_1$ 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.[23] $\theta_3$ measures the vertical distance between the upper and lower

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

[23] Since the arctan function is strictly decreasing, there is no reserve range where curve (8) is perfectly flat, which would correspond to the abundant-reserve region. For inference on the transition between abundant

asymptotes of the nonlinear time-invariant function in (8): $\lim_{x\to-\infty} f(x;\theta) - \lim_{x\to+\infty} f(x;\theta) = \pi\theta_3$. 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.[24] As a result, $\theta_3$ should be (at least) of the same order of magnitude as the DW-IORB spread.

To ensure that minimizing (9) leads to economically meaningful results, we use the framework of Section 2.1 to initialize the parameters and set bounds on them. Appendix D 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 (9) 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.

## 6.3 Results

Figure 10 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 (6). The estimates of the shifts are in Table 2; the estimates of the parameters governing the nonlinear shape of the curve are in Table 3.

The reserve demand curve has moved vertically and horizontally over time. As shown in Table 2, from 2010-2014 to 2015-2020, 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 4 and suggest that, although there seems to

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.

[24] Stigma or borrowing caps may also push the federal funds rate above the DW rate.

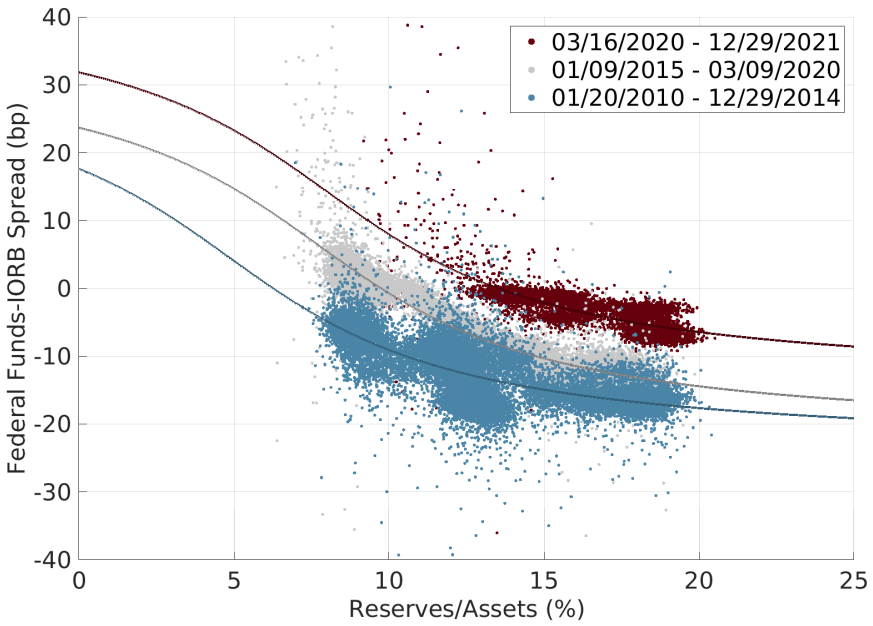


Figure 10: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data.** This figure shows the results of the nonlinear least-squares (NLLS) minimization in equation (9). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of the time-varying VAR model in (6), 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 (8). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix D.

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.

In terms of the time-invariant nonlinear part of the reserve demand function, our estimates show that the point of maximum slope ($\theta_1$) 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 2015-2020, 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 $\theta_3$ 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 $\pi$ (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 | 2.85 | 3.15 |
| Vertical shifts $p^*$ (bp) | -24 | -22.11 | -14.28 |

Table 2: **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 (8) from the nonlinear least-squares (NLLS) minimization in equation (9). 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 (8). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix D.

| Parameter | $\theta_1$ (%) | $\theta_2$ (%) | $\theta_3$ (bp) |
|---|---|---|---|
| | 4.82 | 5.55 | 18.22 |

Table 3: **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 (8) from the nonlinear least-squares (NLLS) minimization in equation (9). 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 (8). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix D.

In Appendix D, for robustness, we fit the demand curve (8) over a broader region that includes the point of maximum slope. Results are similar.

## 6.4 Possible drivers of the low-frequency shifts

What drives the vertical movements of the curve in our sample? An important driver could be the supplementary leverage ratio (SLR) introduced by the Basel III regulation after the 2008 crisis. By constraining leverage, this regulation penalizes balance-sheet expansions and limits banks' ability to run arbitrage strategies, such as the IORB arbitrage. Moreover, because this capital ratio is unweighted, it particularly penalizes balance-sheet expansions that involve investing in safe assets such as reserves (Duffie, 2018). The adoption of the SLR in the US was announced in December 2011 and became effective in April 2014.

Another balance-sheet cost is the FDIC assessment fee, which banks must pay to fund the FDIC's Deposit Insurance Fund. Before 2020, this fee was proportional to a bank's total domestic deposits; since 2011, it has been proportional to the bank's average consolidated total assets minus its average tangible equity. That is, banks pay this fee on their total liabilities. Both the SLR and the revised FDIC fee, therefore, increased banks' balance-sheet costs in the earlier part of our sample, placing downward pressure on $p^*$.

Banks and other market participants respond to regulation and try to minimize regulatory costs. Any market innovation that allows for balance-sheet netting—e.g., through a Central Counterparty Clearing House (CCP)—would reduce banks' balance-sheet costs, putting upward pressure on $p^*$. An important example is the FICC sponsored repo pro-

gram, which expands access to the FICC-cleared interdealer repo market to nondealers such as MMFs (cash lenders) and hedge funds (cash borrowers). This allows banks to net their repo positions with a wider set of counterparties, reducing the balance-sheet costs associated with repo intermediation (Afonso et al., 2021). The program started in 2005 but remained small until 2017, when it became available not only to cash lenders but also borrowers; the program expanded further in April 2019, when it allowed a wider set of FICC members to sponsor counterparties into the program. By freeing banks' balance-sheet capacity, these market innovations may have contributed to the increase in $p^*$ we observe in 2015-2019.

Regulation also changes over time in response to stress episodes. Between March 2020 and March 2021, regulatory agencies temporarily modified the SLR to ease strains in the Treasury market due to the Covid crisis and to promote lending to households and businesses. Under this relief program, Treasuries and reserves were excluded from the SLR calculation, making the regulatory constraint less tight (Federal Register, 2020). This temporary regulatory relief may have caused the significant upward shift of the reserve demand curve during the last part of the sample.

Variation in FHLBs' opportunity cost of lending in the federal funds market also moves the reserve demand curve vertically, by affecting FHLBs' reservation price and therefore their bargaining power. An important example of such variation was the introduction of the ONRRP in September 2013, which has provided FHLBs with a new outside option for safe overnight investment. Moreover, the spread between the IORB and ONRRP rates was reduced three times during 2018-2019, going from 25 bp in early June 2018 to 10 bp in May 2019. These structural changes have likely pushed the reserve demand curve upward in the second half of the sample.

Horizontal shifts in the curve are driven by factors that affect banks' demand for reserves for every price level. The liquidity regulation and supervisory stress tests implemented after 2008 provide some examples. The Liquidity Coverage Ratio, introduced in 2015, requires that banks hold enough high-quality liquid assets, such as reserves or Treasuries, to survive a stress scenario lasting for one month. Under the enhanced prudential standards introduced in 2014 by Regulation YY, large financial firms are required to hold liquidity buffers to cover outflows on the first day of a stress-test scenario, without reliance on the Federal Reserve. The demand for reserves may have also increased due to the requirement, introduced by the Dodd-Frank Act of 2011 and implemented over 2012-2013, that large financial firms prepare plans ("living wills") describing their potential orderly resolution. These regulatory and supervisory changes have likely pushed the reserve demand curve to the right, consistent with our evidence of a moderate horizontal shift of 3 pp in 2015-2019 relative to 2010-2014.

Since 2008, the federal funds market has also become less liquid. After reserves became abundant, trading incentives declined, and average daily volume dropped from a pre-crisis level of $220 billion to $70 billion. This has likely pushed banks' precautionary demand for reserves outward. Indeed, survey data collected by the Federal Reserve since 2018 show that banks identify meeting large intraday payments as a major driver of their reserve demand.[25]

The shift of the demand curve to the right need not be permanent, however. First, the banking system responds to liquidity regulation and supervision by improving and optimizing its reserve management strategies. Second, regulations change; the tailoring rules of 2019, for instance, reduced the liquidity and supervisory requirements for smaller banks (Federal Register, 2019). Moreover, the Federal Reserve has recently tried to reduce banks' reluctance (due to stigma, for example) to borrow from the DW in times of stress (Carlson and Rose, 2017). These changes may in time push the demand for reserves back to the left.

# 7 Conclusion

In this paper, we study the U.S. banking system's 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 12 years of data, from 2010 to 2021.

Our methodology uses a time-varying instrumental-variable estimation at the daily frequency, combined with a vector autoregressive model of the joint dynamics of rates and reserves. Our approach addresses the three main issues affecting the estimation of the reserve demand curve: nonlinearity, time variation due to slow-moving structural changes, and endogeneity. We also show that our methodology works well out-of-sample and can be used as tool to monitor reserve ampleness in real time.

In addition, we provide three main empirical findings that are relevant for understanding the properties of the demand for reserves and the implementation of monetary policy. First, as predicted by economic theory, we find that the reserve demand curve is highly nonlinear with a clear satiation level: it is flat when reserves in the banking system 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 banks' assets; in the second half, when reserves drop below 13%. These findings suggest that the reserve demand

[25] See `https://www.federalreserve.gov/data/sfos/sfos-release-dates.htm`.

curve, and hence its satiation point, have moved outward over time. Third, we show that the curve has also shifted vertically. Upward vertical shifts seem to be 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. While our focus is on assessing the demand-satiation point and documenting shifts of the demand function rather than explaining them, studying the causes of the shifts we document is an important task for future research.

# 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. (2022). Liquidity, liquidity everywhere, not a drop to use - why flooding banks with central bank reserves may not expand liquidity. *NBER Working Paper*, 29680.

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. (2022). Banks' balance-sheet costs, monetary policy, and the ON RRP. *Federal Reserve Bank of New York Staff Reports No. 1041*.

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. 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*.

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.

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*.

Carlson, M. and Rose, J. D. (2017). Stigma and the discount window. `https://www.federalreserve.gov/econres/notes/feds-notes/stigma-and-the-discount-window-20171219.html`.

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. (2021). Reserves were not so ample after all. *Federal Reserve Bank of New York Staff Reports No. 974*.

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. (2020). The reserve supply channel of unconventional monetary policy. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Duffie, D. (2018). Post-crisis bank regulations and financial market liquidity. `https://www.darrellduffie.com/uploads/policy/DuffieBaffiLecture2018.pdf`.

d'Avernas, A. and Vandeweyer, Q. (2021). Intraday liquidity and money market dislocations. *Working Paper*.

d'Avernas, A. and Vandeweyer, Q. (2023). Treasury bill shortages and the pricing of short-term assets. *Journal of Finance (forthcoming)*.

Ennis, H. M. and Keister, T. (2008). Understanding monetary policy implementation. *Economic Quarterly*, 94:235–263.

Federal Register (2019). Company-run stress testing requirements for FDIC-supervised state nonmember banks and state savings associations.

Federal Register (2020). Temporary exclusion of U.S. Treasury securities and deposits at federal reserve banks from the supplementary leverage ratio.

FOMC (2019). Statement regarding monetary policy implementation. `https://www.federalreserve.gov/newsevents/pressreleases/monetary20191011a.htm`.

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.

Kashyap, A. and Stein, J. C. (2012). The optimal conduct of monetary policy with interest on reserves. *American Economic Journal: Macroeconomics*, 4:266–282.

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.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

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.

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.

# Internet Appendix

## Appendix A Demand for Reserves with 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 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.[26] 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 $\kappa$ is the same for all banks.

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

$$\min_{f_i} \left[ (i^f - i^{IORB} + \kappa) \int_{\hat{z}_i}^{\overline{z}} (r_i' - \bar{r}_i) g(z) dz + (i^{DW} - i^f) \int_{\underline{z}}^{\hat{z}_i} (\bar{r}_i - r_i') g(z) dz - \kappa \min\{0, f_i\} \right], \tag{A.1}$$

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

Objective function (A.1) is not differentiable due to the minimum function in the last term. To simplify the derivation, we use a smooth approximation of the minimum: $\min\{0, x\} \approx -\frac{1}{\alpha} \ln\left(1 + e^{-\alpha x}\right)$ 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.1) to derive the optimal federal funds net borrowing:

$$i^f = (i^{IORB} - \kappa) + (i^{DW} - i^{IORB} + \kappa)\, G(\bar{r}_i - (\widetilde{r}_i + f_i^*)) + \frac{\kappa}{1 + e^{\alpha f_i^*}}, \tag{A.2}$$

where $f_i^*$ is the unique minimizer of equation (A.1).

Equation (A.2) represents the inverse demand for reserves of an individual bank in the

[26] We can easily generalize this assumption to include loans and other assets.

presence of (smooth) balance-sheet costs proportional to balance-sheet size. It is easy to see that the $i^f$ 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. Since the aggregate inverse demand for reserves is the horizontal summation of the individual curves, it retains these properties, as depicted in Figure 1 in the text.

# 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:

$$\begin{aligned} q_t &= c_{q,t} + b_{q,q,1,t} q_{t-1} + b_{q,p,1,t} p_{t-1} + ... + b_{q,q,m,t} q_{t-m} + b_{q,p,m,t} s_{t-m} + u_{q,t}, \\ p_t &= c_{p,t} + b_{p,q,1,t} q_{t-1} + b_{p,p,1,t} p_{t-1} + ... + b_{p,q,m,t} q_{t-m} + b_{p,p,m,t} s_{t-m} + u_{s,t}, \end{aligned} \tag{A.3}$$

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 $2 \times 2$ covariance matrix $\Omega_t$; 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.3) is:

$$y_t = c_t + B_{1,t} y_{t-1} + ... + B_{m,t} y_{t-m} + u_t \qquad \text{with } t = 1, ..., T, \tag{A.4}$$

where $y_t$ is a $2 \times 1$ stacked vector of $(q_t, p_t)'$; $c_t$ is an $2 \times 1$ vector of stacked constant terms $(c_{q,t}, c_{p,t})'$; $B_{i,t}$, with $i = 1, \ldots, m$, are the following $2 \times 2$ 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 $\Omega_t$ as

follows:

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \tag{A.5}$$

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

$$\begin{aligned} y_t &= c_t + B_{1,t} y_{t-1} + ... + B_{m,t} y_{t-m} + A_t^{-1} \Sigma_t \varepsilon_t, \\ Var(\varepsilon_t) &= I_n, \end{aligned} \tag{A.6}$$

where $\varepsilon_t$ is a $2 \times 1$ 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.5) is convenient because the first $\varepsilon$ 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:

$$\begin{aligned} y_t &= X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \\ X_t' &= I_n \otimes [1, y_{t-1}', \ldots, y_{t-m}'], \end{aligned} \tag{A.7}$$

where $\otimes$ denotes the Kronecker product.

We model the parameters in the following way:

$$B_t = B_{t-1} + \nu_t, \tag{A.8}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t, \tag{A.9}$$

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t, \tag{A.10}$$

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 $2 \times 1$ vector of diagonal terms in $\Sigma_t$. $B$ and $\alpha$ are modeled as random walks; $\sigma_t$ 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

$$V = Var\left(\begin{bmatrix}\epsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t\end{bmatrix}\right) = \begin{bmatrix} I_2 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \tag{A.11}$$

where $I_2$ is the $2 \times 2$ identity matrix, $S$ is the variance of $\zeta_t$, 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 $3 \times 1$ 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 $\alpha_t$ in (A.9) becomes a $3 \times 1$ vector of the stacked parameters of $A_t$. $\Sigma_t$ maintains its diagonal structure and expands to include $\sigma_{3,t}$, so that $\sigma_t$ in (A.10) becomes a $3 \times 1$ vector.

The covariance matrix of $\varepsilon_t$ 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 $\zeta$ innovation for $\alpha_{21}$, and the lower block is the covariance of the $\zeta$ innovations for $(\alpha_{31}, \alpha_{32})'$.

## Appendix B.2 Covariance between errors and observables: an impulse-response view

In this section, we show how the covariances in equation (7) can be interpreted as the $h$-day-ahead impulse responses of rates and reserves to a reserve shock under a Choleski decomposition with reserves ordered first, such as the factorization in (A.5).

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 $\Gamma_{1:T} = \{c_t, B_{1,t}, ..., B_{m,t}, A_t, \Sigma_t;\ t = 1, ..., T\}$. 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

$$\begin{aligned}\operatorname{cov}\left(y_t, u_{1,t-h}|\Gamma_{1:T}\right) &= \mathrm{E}\left[y_t u_{1,t-h}|\Gamma_{1:T}\right] = \mathbf{J}'\prod_{k=1}^{h}\mathbf{B}_{t-k}\mathrm{E}\left[\mathbf{u}_{t-h}\mathbf{u}_{t-h}'|\Gamma_{1:T}\right]\mathbf{j}_1 = \mathbf{J}'\prod_{k=1}^{h}\mathbf{B}_{t-k}\mathbf{J}\Omega_{t-h}\mathbf{J}'\mathbf{j}_1, \\ &= \mathbf{J}'\prod_{k=1}^{h}\mathbf{B}_{t-k}\mathbf{J}\Omega_{t-h}\iota_1\end{aligned}$$

where $\iota_1 = (1 \underbrace{0 \ldots 0}_{n-1 \text{ times}})'$. $\Omega_{t-h}\iota_1$ is the first column of the covariance matrix $\Omega_{t-h}$.

The Cholesky factorization (A.5) implies that $\Omega_t\iota_1 = A_t^{-1}\Sigma_t\Sigma_t'\left(A_t'\right)^{-1}\iota_1 = \left(A_t^{-1}\iota_1\right)\sigma_{1t}^2$. As a result,

$$\operatorname{cov}\left(y_t, u_{1,t-h}|\Gamma_{1:T}\right) = \left(\mathbf{J}'\prod_{k=1}^{h}\mathbf{B}_{t-k}\mathbf{J}A_{t-h}^{-1}\iota_1\right)\sigma_{1t-h}^2. \tag{A.12}$$

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.12) are therefore equal to the $h$-day-ahead impulse responses of the system variables to the standardized reserve shocks calculated using factorization (A.5); in fact, in the traditional VAR literature, the $i$-th variable's impulse response to $\varepsilon_1$ after $h$ days, $\dfrac{\partial y_{i,t+h}}{\partial \varepsilon_{1,t}}$, 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 (7), obtained as ratio of the covariances in (A.12), is exactly equal to the ratio of the $h$-day-ahead impulse responses of rates and reserves to reserve shocks.

## Appendix B.3 Priors

We use Bayesian methods to estimate model (A.3). 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 $\mu$ and variance $\sigma^2$, 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.5) 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 $\lambda_i$ 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).

## Appendix B.4 Out-of-sample validation: forecasting

As discussed in Section 4.3, 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.3), 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 29, 2021, 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.

### Appendix B.4.1 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 4 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 period, when reserves were so abundant that rate fluctuations were unrelated to reserve shocks, are results more mixed, with no model dominating the others.

### Appendix B.4.2 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

| 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 - 2021 | TV-VAR | 0.413 | 0.606 | 0.876 | 1.67 | 2.03 | 2.31 | 0.677 | 1.20 | 1.97 |
| | VAR | 0.439 | 0.653 | 1.02 | 1.99 | 2.03 | 2.64 | 0.854 | 1.28 | 2.44 |
| | AR | 0.426 | 0.639 | 1.04 | 2.66 | 3.87 | 6.18 | - | - | - |
| 2010 - 2011 | TV-VAR | 0.343 | 0.457 | 0.679 | 1.47 | 1.76 | 2.24 | 0.444 | 0.737 | 1.44 |
| | VAR | 0.378 | 0.528 | 0.807 | 1.66 | 2.09 | 2.93 | 0.521 | 0.902 | 1.81 |
| | AR | 0.359 | 0.503 | 0.747 | 1.58 | 1.99 | 2.77 | - | - | - |
| 2012 - 2013 | TV-VAR | 0.340 | 0.440 | 0.534 | 1.27 | 1.74 | 2.45 | 0.419 | 0.725 | 1.22 |
| | VAR | 0.347 | 0.474 | 0.654 | 1.38 | 1.88 | 2.70 | 0.456 | 0.801 | 1.49 |
| | AR | 0.347 | 0.456 | 0.612 | 1.37 | 1.90 | 2.69 | - | - | - |
| 2014 - 2015 | TV-VAR | 0.584 | 0.834 | 1.05 | 0.778 | 1.03 | 1.27 | 0.454 | 0.863 | 1.33 |
| | VAR | 0.617 | 0.773 | 0.815 | 0.820 | 1.04 | 1.28 | 0.506 | 0.806 | 1.04 |
| | AR | 0.587 | 0.767 | 0.849 | 0.778 | 0.986 | 1.29 | - | - | - |
| 2016 - 2017 | TV-VAR | 0.364 | 0.537 | 0.757 | 0.423 | 0.648 | 0.968 | 0.152 | 0.343 | 0.725 |
| | VAR | 0.385 | 0.609 | 0.833 | 0.473 | 0.713 | 0.982 | 0.174 | 0.415 | 0.775 |
| | AR | 0.373 | 0.568 | 0.773 | 0.462 | 0.709 | 1.04 | - | - | - |
| 2018 - 2019 | TV-VAR | 0.263 | 0.370 | 0.475 | 2.44 | 2.77 | 2.97 | 0.558 | 0.818 | 1.13 |
| | VAR | 0.312 | 0.454 | 0.543 | 3.35 | 2.70 | 3.52 | 0.927 | 1.04 | 1.69 |
| | AR | 0.270 | 0.391 | 0.522 | 5.43 | 8.28 | 14.0 | - | - | - |
| 2020 - 2021 | TV-VAR | 0.490 | 0.810 | 1.39 | 2.48 | 3.06 | 3.10 | 1.21 | 2.43 | 4.18 |
| | VAR | 0.513 | 0.926 | 1.87 | 2.58 | 2.85 | 3.33 | 1.32 | 2.60 | 5.28 |
| | AR | 0.526 | 0.950 | 1.99 | 2.67 | 3.26 | 3.36 | - | - | - |

Table 4: **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.3) and the VAR and AR models described in Appendix B.4.

the forecasting model. Lower scores correspond to lower likelihoods, suggesting lower model accuracy.

Table 5 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 10 and 20 business days, though it is only materially lower for the latter. Moreover, when considering the performance by sub-period, all log-scores from the TV-VAR, including those for the spread, 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 part (2014-2017), 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 on $[0, 1]$. Figure 11 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-degrees 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.

| 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 - 2021 | TV-VAR | -0.509 | -0.863 | -1.38 | -2.06 | -2.39 | -3.33 | -2.60 | -3.26 | -4.68 |
| | VAR | -0.682 | -1.17 | -1.97 | -2.19 | -2.27 | -2.66 | -3.10 | -3.70 | -4.98 |
| | AR | -0.600 | -1.08 | -1.83 | -2.19 | -2.33 | -2.59 | - | - | - |
| 2010 - 2011 | TV-VAR | -0.366 | -0.655 | -1.13 | -1.82 | -2.02 | -2.29 | -2.07 | -2.63 | -3.36 |
| | VAR | -0.574 | -0.957 | -1.76 | -2.02 | -2.29 | -2.76 | -2.40 | -3.09 | -4.44 |
| | AR | -0.493 | -0.861 | -1.47 | -1.94 | -2.19 | -2.60 | - | - | - |
| 2012 - 2013 | TV-VAR | -0.368 | -0.661 | -0.966 | -1.86 | -2.20 | -2.67 | -2.23 | -2.90 | -3.62 |
| | VAR | -0.429 | -0.825 | -1.31 | -1.75 | -2.13 | -2.70 | -2.13 | -2.88 | -3.84 |
| | AR | -0.412 | -0.738 | -1.09 | -1.74 | -2.14 | -2.67 | - | - | - |
| 2014 - 2015 | TV-VAR | -1.11 | -1.43 | -1.62 | -1.62 | -2.46 | -4.28 | -2.74 | -3.93 | -5.85 |
| | VAR | -1.25 | -1.40 | -1.42 | -1.30 | -1.59 | -1.90 | -2.63 | -3.16 | -3.42 |
| | AR | -1.10 | -1.39 | -1.49 | -1.21 | -1.47 | -1.86 | - | - | - |
| 2016 - 2017 | TV-VAR | -0.432 | -0.838 | -1.55 | -1.09 | -2.25 | -4.24 | -1.52 | -3.08 | -5.76 |
| | VAR | -0.463 | -0.945 | -1.32 | -0.551 | -0.901 | -1.31 | -1.04 | -1.93 | -2.82 |
| | AR | -0.425 | -0.833 | -1.16 | -0.547 | -0.890 | -1.31 | - | - | - |
| 2018 - 2019 | TV-VAR | -0.166 | -0.536 | -0.967 | -4.44 | -3.23 | -3.31 | -4.84 | -3.74 | -4.24 |
| | VAR | -0.346 | -0.763 | -0.945 | -5.56 | -4.55 | -4.75 | -6.93 | -6.16 | -6.64 |
| | AR | -0.128 | -0.516 | -0.862 | -5.65 | -4.90 | -4.46 | - | - | - |
| 2020 - 2021 | TV-VAR | -0.596 | -1.02 | -1.99 | -1.50 | -2.11 | -3.07 | -2.12 | -3.20 | -5.07 |
| | VAR | -1.02 | -2.08 | -5.03 | -1.93 | -2.17 | -2.54 | -3.34 | -4.92 | -8.61 |
| | AR | -1.02 | -2.10 | -4.84 | -1.99 | -2.36 | -2.66 | - | - | - |

Table 5: **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.3) and the VAR and AR models described in Appendix B.4.

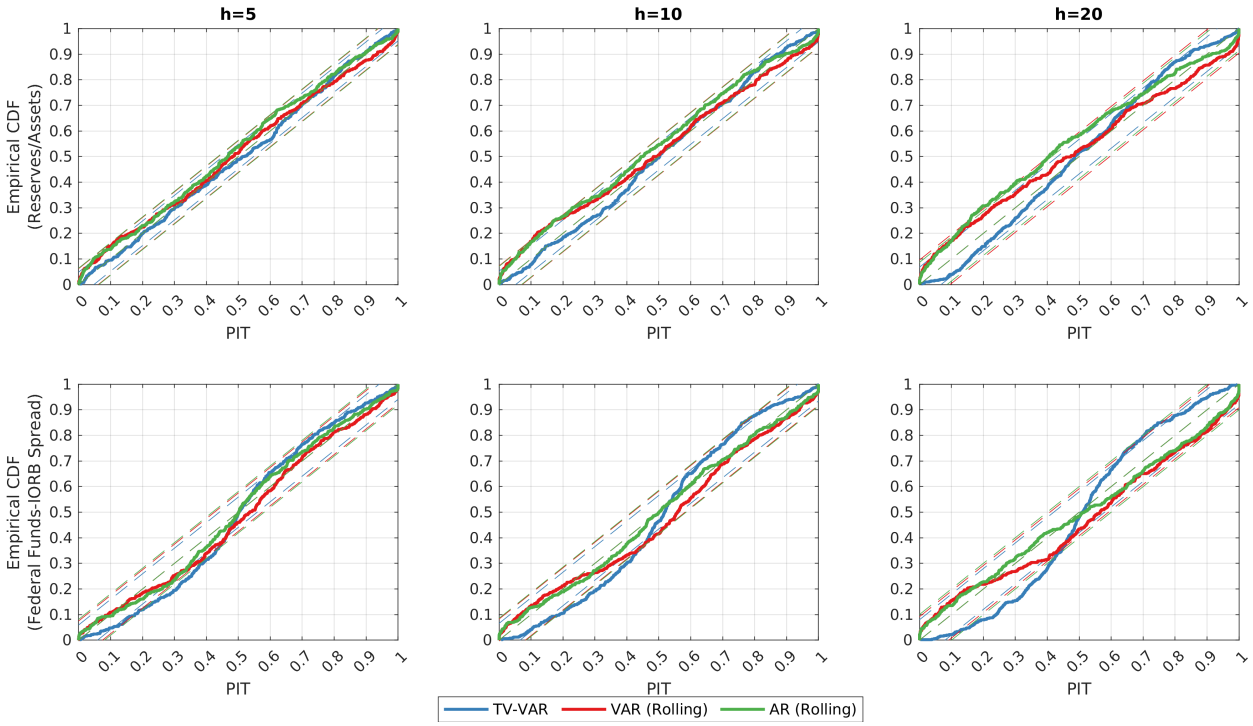


Figure 11: **Empirical cumulative distribution function (CDF) of the probability integral transforms (PITs)**. Empirical CDFs of the PITs of the forecasts of normalized reserves and of the federal-funds-IORB spread from the TV-VAR (A.3) and the VAR and AR models described in Appendix B.4. The colored dashed lines represent the 90% confidence bands for the different models, constructed using the bootstrap method outlined in Rossi and Sekhposyan (2019).

# Appendix C Different Normalizations of Reserves

## Appendix C.1 Normalization of Reserves by 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 12 shows the evolution of reserves normalized by deposits over 2010-2021. Compared to reserves normalized by banks' assets, the level is higher (as banks' deposits are smaller than banks' assets), but the time variation is almost identical.

Figure 13 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 (6); Figures 14, and 15 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 6, 7, and 8): 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 6 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 0.23 bp in 2010 and 0.18 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).

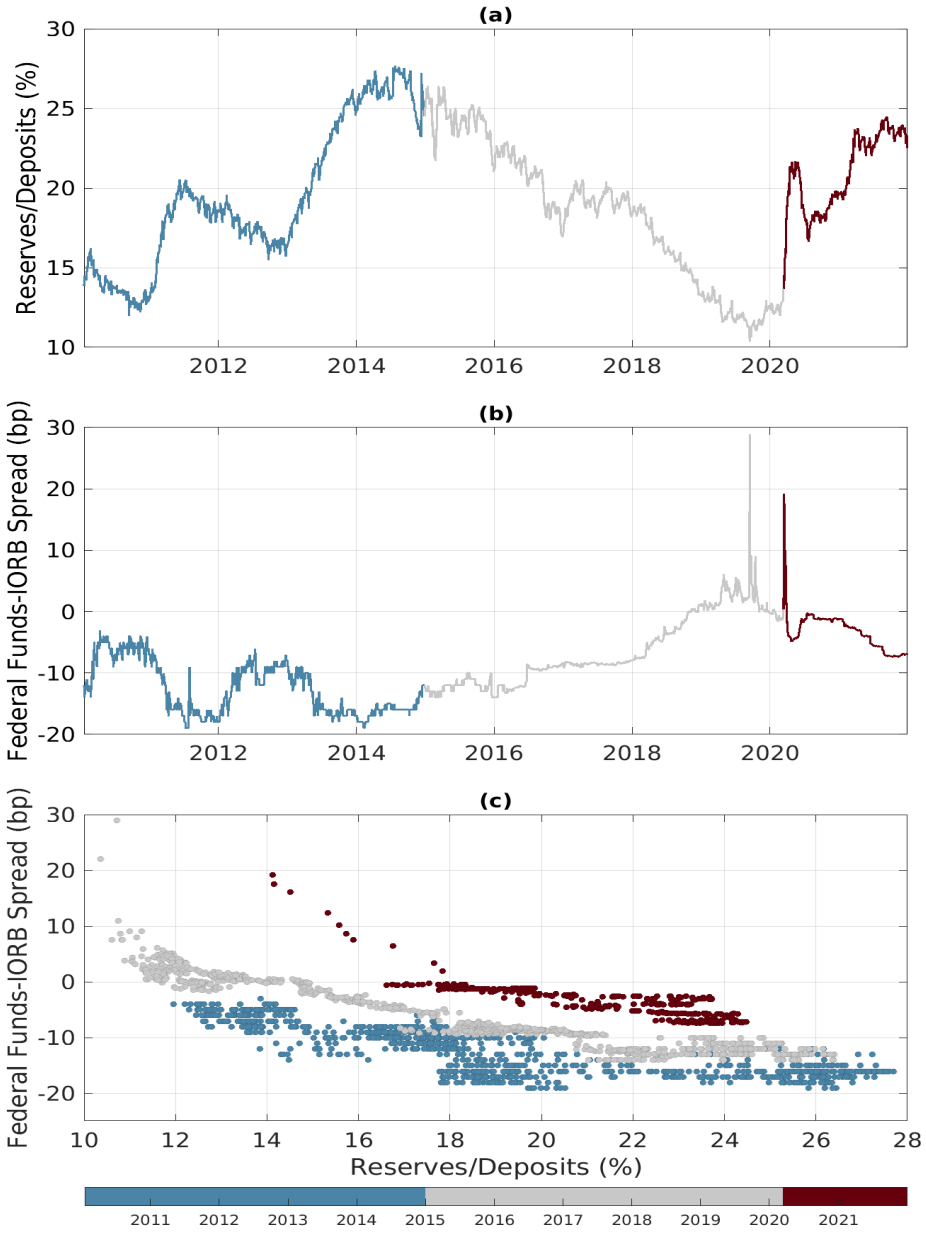


Figure 12: **Deposit-normalized reserves, the federal funds rate, and the reserve demand curve.** Panel (a) plots aggregate reserves relative to to commercial banks' deposits from January 1, 2010 to December 29, 2021. Panel (b) shows the spread between the volume-weighted average federal funds rate and the IORB rate (in basis points). Panel (c) plots the relationship between the spread and normalized reserves. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. 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"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

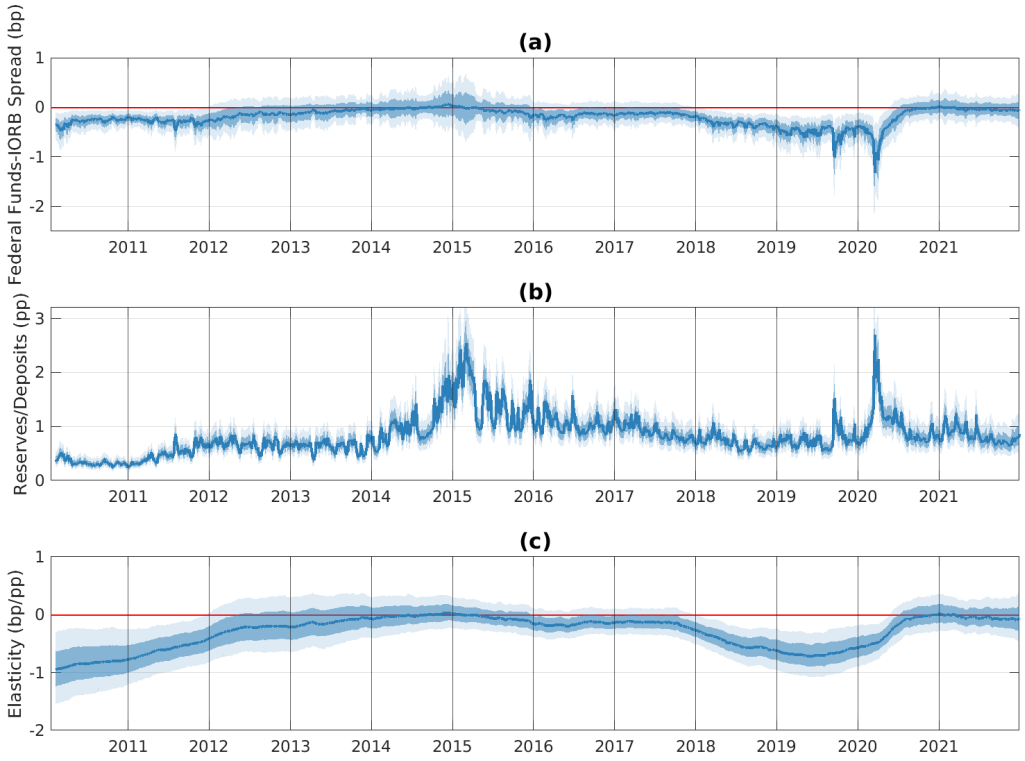


Figure 13: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample from model (6) 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 deposits. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. 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"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

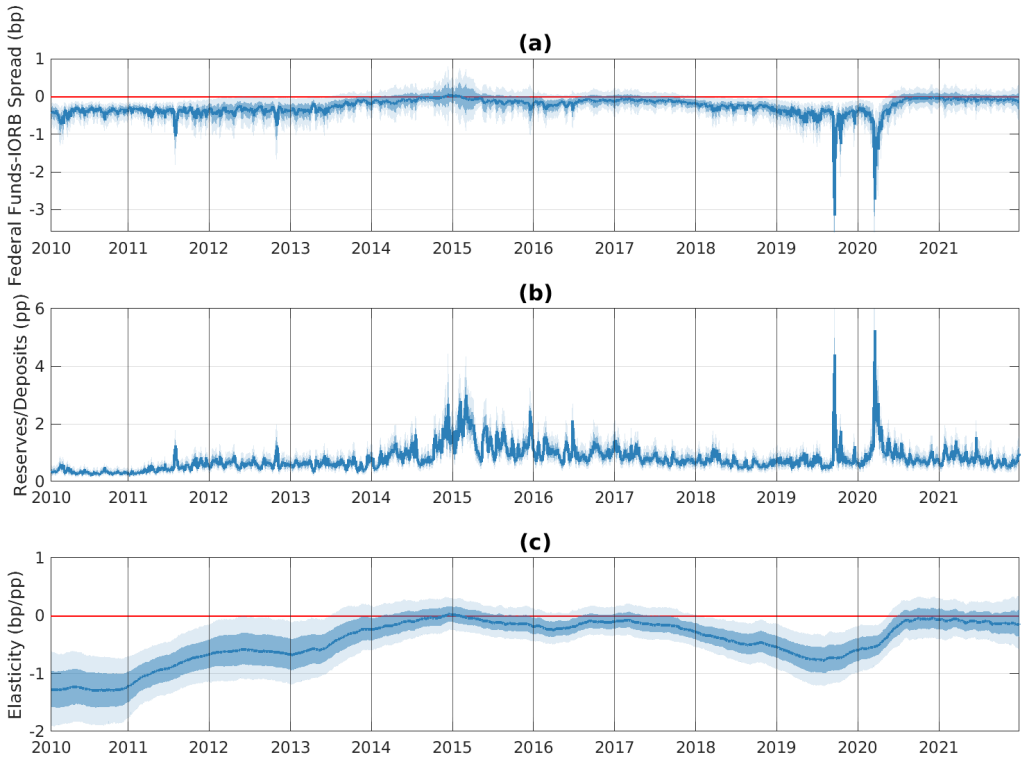


Figure 14: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves controlling for repo rates.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (6) 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 deposits. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York; daily data on repo rates are for overnight Treasury repos and available from the Depository Trust & Clearing Corporation (DTCC) at `https://www.dtcc.com/charts/dtcc-gcf-repo-index`. 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"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

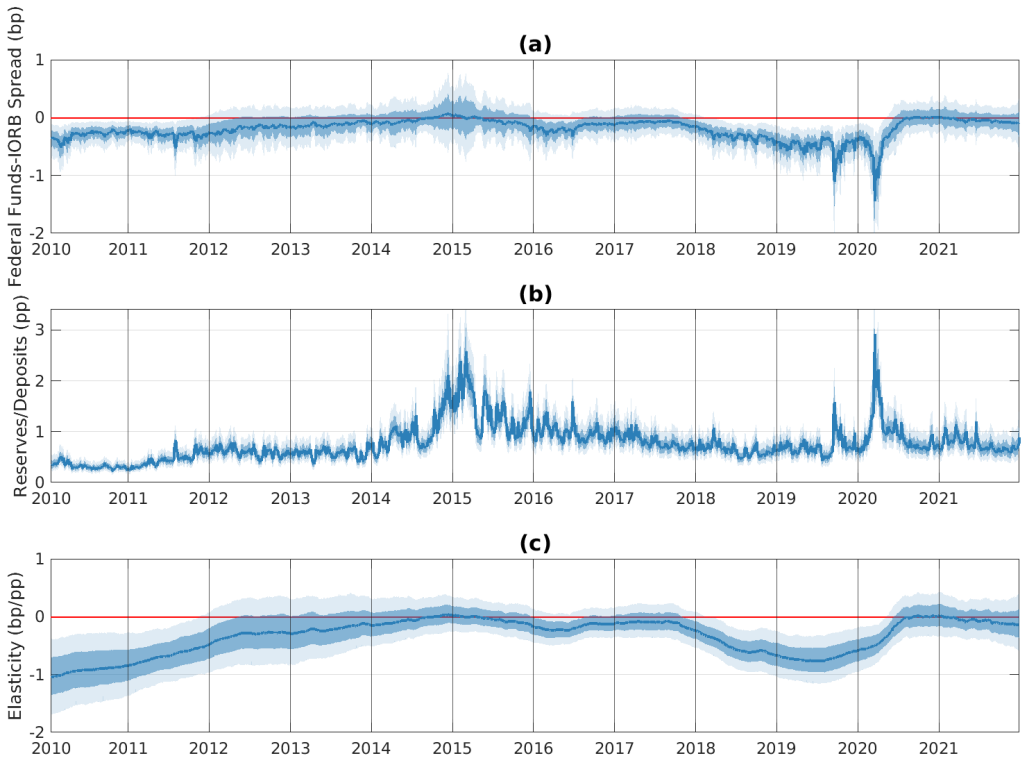


Figure 15: **IV estimate of the elasticity of the federal funds rate to deposit-normalized reserves controlling for Treasury yields.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (6) 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 deposits. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. 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"). The daily interest rate on reserve balances and daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER", "IORB", and "DGS1" respectively).

| 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (a) Bi-variate Model | | | | | | | | | | | |
| -0.84 | -0.59 | -0.24 | -0.12 | -0.01 | -0.04 | -0.15 | -0.14 | -0.47 | -0.66 | -0.21 | -0.04 |
| (-1.39,-0.25) | (-1.05,-0.13) | (-0.69,0.26) | (-0.54,0.34) | (-0.32,0.32) | (-0.3,0.23) | (-0.39,0.1) | (-0.39,0.1) | (-0.84,-0.12) | (-1.03,-0.26) | (-0.74,0.24) | (-0.35,0.29) |
| (b) Tri-variate Model with Repo Rates | | | | | | | | | | | |
| -1.26 | -0.9 | -0.62 | -0.45 | -0.09 | -0.09 | -0.16 | -0.15 | -0.42 | -0.68 | -0.23 | -0.1 |
| (-1.87,-0.68) | (-1.47,-0.34) | (-1.12,-0.05) | (-1.03,0.09) | (-0.46,0.26) | (-0.37,0.21) | (-0.43,0.12) | (-0.43,0.15) | (-0.78,-0.07) | (-1.13,-0.26) | (-0.8,0.24) | (-0.44,0.26) |
| (c) Tri-variate Model with Treasury Yields | | | | | | | | | | | |
| -0.92 | -0.66 | -0.31 | -0.2 | -0.05 | -0.04 | -0.17 | -0.11 | -0.49 | -0.71 | -0.19 | -0.06 |
| (-1.53,-0.3) | (-1.18,-0.12) | (-0.86,0.29) | (-0.69,0.35) | (-0.41,0.34) | (-0.34,0.27) | (-0.45,0.11) | (-0.4,0.2) | (-0.91,-0.07) | (-1.11,-0.28) | (-0.77,0.32) | (-0.43,0.32) |



Table 6: **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 (7). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (6); in panel (b), they are estimated from an augmented trivariate version of model (6) that also includes daily repo rates; in panel (c), they are estimated from an augmented trivariate version of model (6) that includes daily Treasury yields. All multivariate models include ten lags ($m = 10$ days). The reported elasticities are calculated by year. Reserves are measured, in percent, as a ratio to deposits. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York; daily data on repo rates are for overnight Treasury repos and available from the Depository Trust & Clearing Corporation (DTCC) at `https://www.dtcc.com/charts/dtcc-gcf-repo-index`. 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"). The daily interest rate on reserve balances and the daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER", "IORB", and "DGS1" respectively).

## Appendix C.2 Normalization of Reserves 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 data on U.S. GDP are from FRED; we linearly interpolate these quarterly observations to obtain daily estimates. Figure 16 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 17 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 (6); Figures 18, and 19 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 6, 7, and 8): 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 7 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 1.6 bp in 2011 and by 1.1 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.

# Appendix D The Post-processing Nonlinear Fit

## Appendix D.1 Algorithm details

To solve the minimization problem in (9), 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 (9) 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

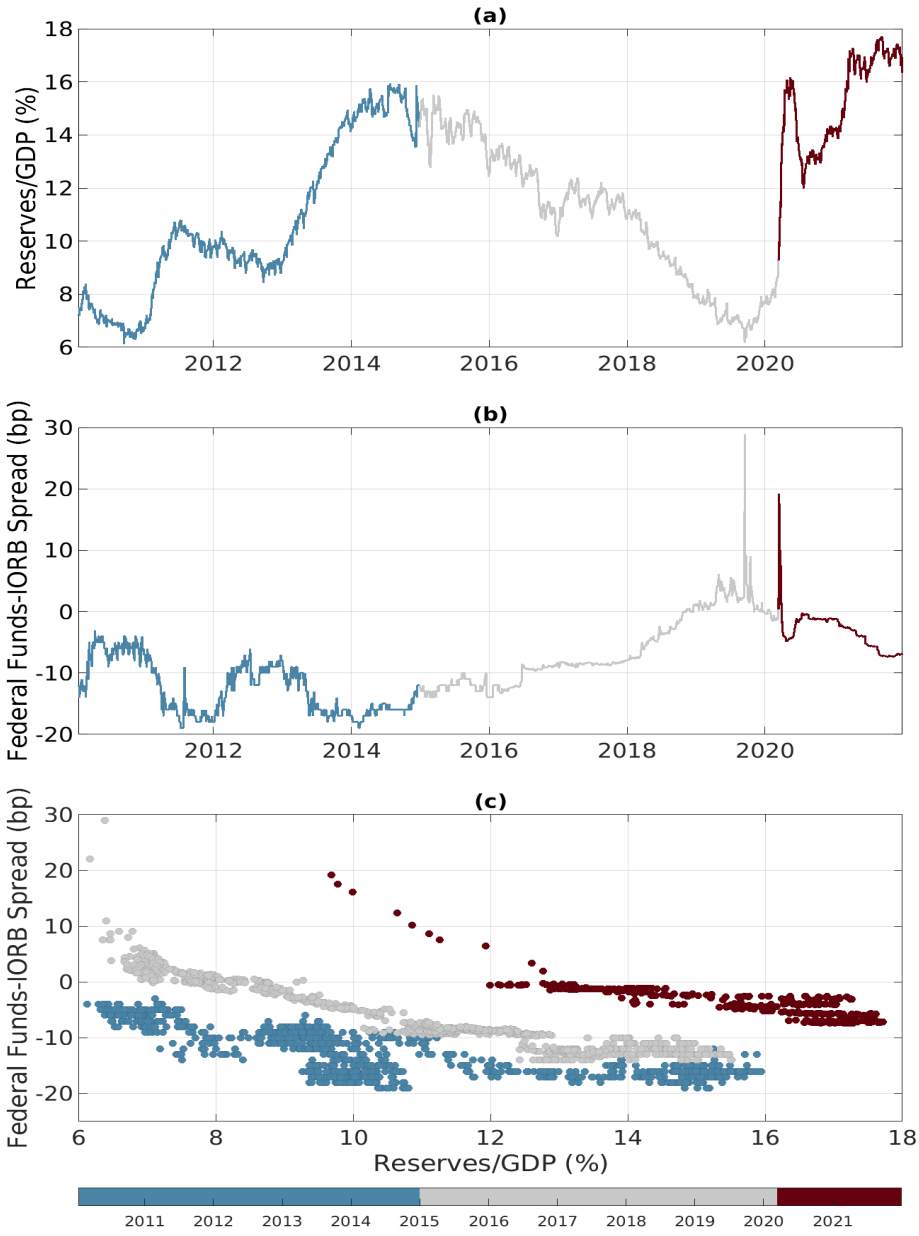


Figure 16: **GDP-normalized reserves, the federal funds rate, and the reserve demand curve.** Panel (a) plots aggregate reserves relative to GDP from January 1, 2010 to December 29, 2021. Panel (b) shows the spread between the volume-weighted average federal funds rate and the IORB rate (in basis points). Panel (c) plots the relationship between the spread and normalized reserves. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Quarterly data on GDP are publicly available from the Federal Reserve Economic Data, FRED ("GDP"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

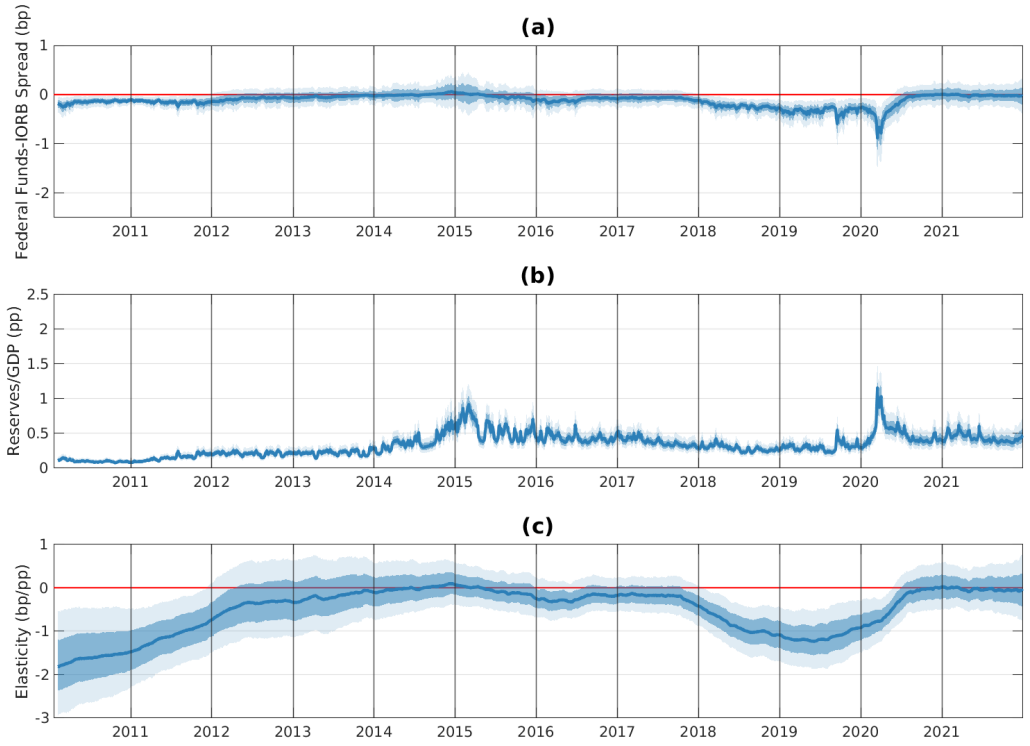


Figure 17: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample from model (6) 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. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Quarterly data on GDP are publicly available from the Federal Reserve Economic Data, FRED ("GDP"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

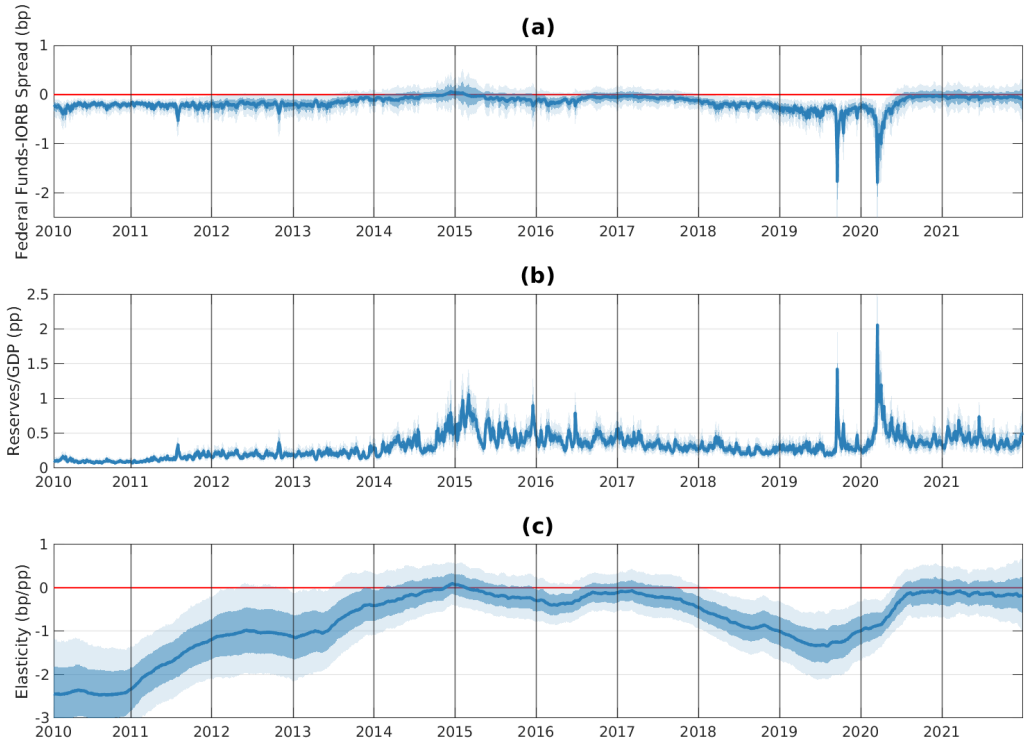


Figure 18: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves controlling for repo rates.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (6) 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. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York; daily data on repo rates are for overnight Treasury repos and available from the Depository Trust & Clearing Corporation (DTCC) at `https://www.dtcc.com/charts/dtcc-gcf-repo-index`. Quarterly data on GDP are publicly available from the Federal Reserve Economic Data, FRED ("GDP"). The daily interest rate on reserve balances is available from FRED ("IOER" and "IORB").

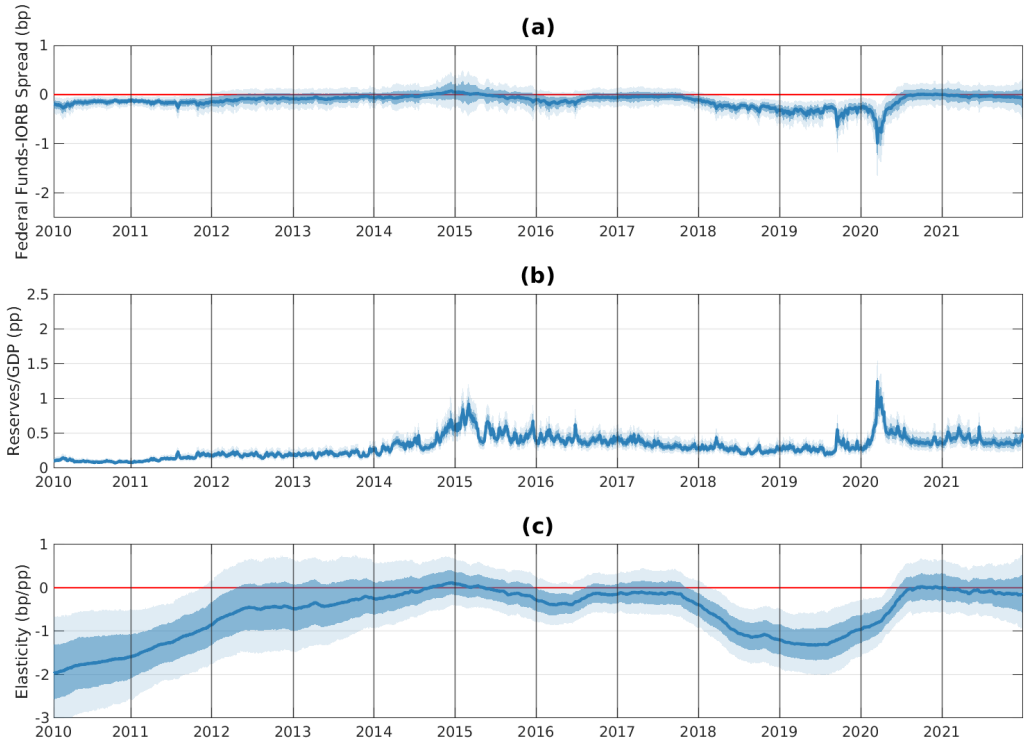


Figure 19: **IV estimate of the elasticity of the federal funds rate to GDP-normalized reserves controlling for Treasury yields.** The IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of the federal funds rate (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (7). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (6) 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. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York. Quarterly data on GDP are publicly available from the Federal Reserve Economic Data, FRED ("GDP"). The daily interest rate on reserve balances and daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER", "IORB", and "DGS1" respectively).

| 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (a) Bi-variate Model | | | | | | | | | | | |
| -1.61 | -1.1 | -0.41 | -0.19 | 0 | -0.07 | -0.24 | -0.2 | -0.84 | -1.13 | -0.34 | -0.04 |
| (-2.66,-0.47) | (-1.99,-0.22) | (-1.25,0.55) | (-0.97,0.68) | (-0.58,0.61) | (-0.54,0.43) | (-0.67,0.21) | (-0.64,0.24) | (-1.51,-0.19) | (-1.78,-0.42) | (-1.2,0.4) | (-0.56,0.54) |
| (b) Tri-variate Model with Repo Rates | | | | | | | | | | | |
| -2.43 | -1.69 | -1.06 | -0.78 | -0.15 | -0.14 | -0.25 | -0.21 | -0.78 | -1.2 | -0.39 | -0.14 |
| (-3.6,-1.31) | (-2.81,-0.6) | (-2.03,0.01) | (-1.84,0.2) | (-0.84,0.49) | (-0.65,0.4) | (-0.75,0.25) | (-0.73,0.32) | (-1.44,-0.13) | (-1.97,-0.47) | (-1.32,0.39) | (-0.72,0.48) |
| (c) Tri-variate Model with Treasury Yields | | | | | | | | | | | |
| -1.75 | -1.23 | -0.53 | -0.35 | -0.07 | -0.06 | -0.26 | -0.16 | -0.92 | -1.22 | -0.31 | -0.09 |
| (-2.9,-0.56) | (-2.22,-0.21) | (-1.54,0.59) | (-1.26,0.67) | (-0.75,0.63) | (-0.59,0.5) | (-0.76,0.24) | (-0.68,0.38) | (-1.66,-0.13) | (-1.93,-0.47) | (-1.26,0.5) | (-0.7,0.55) |



Table 7: **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 (7). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (6); in panel (b), they are estimated from an augmented trivariate version of model (6) that also includes daily repo rates; in panel (c), they are estimated from an augmented trivariate version of model (6) that includes daily Treasury yields. All multivariate models include ten lags ($m = 10$ days). The reported elasticities are calculated by year. Reserves are measured, in percent, as a ratio to GDP. The federal funds rate is measured, in basis points, as a spread to the IORB rate. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and the federal funds rate caused by the window-dressing of European banks around month-ends (see Section 2.2). Daily data on reserves and the federal funds rate are collected by the Federal Reserve Bank of New York; daily data on repo rates are for overnight Treasury repos and available from the Depository Trust & Clearing Corporation (DTCC) at `https://www.dtcc.com/charts/dtcc-gcf-repo-index`. Quarterly data on GDP are publicly available from the Federal Reserve Economic Data, FRED ("GDP"). The daily interest rate on reserve balances and the daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER", "IORB", and "DGS1" respectively).

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 Section 6. 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 (8) is strictly positive everywhere, which implies that $p_t^* < p_t$ at all times by construction.

For the $\theta$ parameters governing the nonlinearity of the demand curve in (8), we impose the following bounds. $\theta_1$ represents the point of maximum absolute slope of the demand curve, where reserves are highly scarce. To bound $\theta_1$ 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 $\theta_1$ from above, we multiply the maximum of normalized reserves in our sample by two. We use the same bounds for $\theta_2$, 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 (8).

The upper asymptote of the nonlinear function (8) 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 $\theta_3$ to be equal to the maximum of the spread between the DW and ONRRP rates in our sample divided by $\pi$. The lower bound on $\theta_3$ is simply set to zero.

To initialize the variables, we also build on our discussion of their economic interpretation in Section 6. We initialize $\theta_1$ 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 $\theta_2$, 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 (8) is $\theta_1 + \theta_2/\sqrt{3}$, we initialize $\theta_2$ 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 $\theta_1$. The initial value of $\theta_3$ is set equal to the average spread between the DW and IORB rates in our sample divided by $\pi$. 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 |
|---|---|---|---|
| $\theta_1$ | 0.18 pp | 38.83 pp | 7.30 pp |
| $\theta_2$ | 0.18 pp | 38.83 pp | 2.94 pp |
| $\theta_3$ | 0.00 bp | 25.5 bp | 14.79 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 |
| 03/16/2020-12/29/2021 | -25.00 bp | -3.41 bp | -8.36 bp |

Table 8: **Bounds and initializations of the variables in the nonlinear least squares (NLLS) minimization in equation (9)**. $\theta_1$, $\theta_2$, and $\theta_3$ are the parameters defining the nonlinear time-invariant functional form of the reserve demand function in equation (8); $q^*$ and $p^*$ are the horizontal and vertical shifts.

## Appendix D.2 Robustness

In this section, instead of fitting the curve to the right of the point of maximum slope, we fit it to the right of the point where the curve flattens around its upper asymptote. Namely, we impose $q_{it} - q_t^* > \theta_1 - \theta_2/\sqrt{3}$ because $\theta_1 - \theta_2/\sqrt{3}$ is the point at which the absolute slope of the curve decreases at the fastest rate as reserves decrease (i.e., the curve flattens around its upper asymptote). This constraint is milder than the one used in the main text as this region includes the point of maximum slope.

Results are in Table (9) and are very close to the baseline results in Section 6, confirming that our findings are not driven by our choice of the region over which we fit the demand curve. In particular, from the beginning to the end of the sample, we find a significant upward vertical shift of 10 bp and a more modest horizontal shift to the right of 3 pp.

The estimates of the time-invariant nonlinear curve are also close to the baseline ones: the point of maximum slope is reached when the reserves-to-assets ratio is around 7-8%, and the point of maximum slope growth when this ratio is around 10%. Together with the estimates of the horizontal shift $q^*$, these results suggest that the transition between ample and scarce reserves occurs around 10% of banks' assets in the first half of the sample and around 12% in the second half.

| | 1/2010-12/2014 | 01/2015-03/2020 | 03/2020-12/2021 |
|---|---|---|---|
| Horizontal shifts $q^*$ (pp) | 0 | 2.91 | 3.08 |
| Vertical shifts $p^*$ (bp) | -21.23 | -19.35 | -11.34 |

Table 9: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data - estimates of the shifts when fitting to the right of the curve's upper asymptote.** This table shows the estimates of the horizontal ($q^*$) and vertical ($p^*$) shifts in equation (8) from the nonlinear least-squares (NLLS) minimization in equation (9). 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. Results in panel (a) are obtained using the bivariate model (6) to generate the forecasts; results in panel (b) and (c) are obtained using the trivariate models that add repo rates and Treasury yields to model (6). All models include 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 where the curve flattens around its upper asymptote (i.e., imposing the constraint: $q_{it} - q_t^* > \theta_1 - \theta_2/\sqrt{3}$). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix D.

| Parameter | $\theta_1$ (%) | $\theta_2$ (%) | $\theta_3$ (bp) |
|---|---|---|---|
| | 7.37 | 4.53 | 11.67 |

Table 10: **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 when fitting to the right of the curve's upper asymptote.** This table shows the estimates of $\theta = (\theta_1, \theta_2, \theta_3)$ in equation (8) from the nonlinear least-squares (NLLS) minimization in equation (9). 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. Results in the first row are obtained using the bivariate model (6) to generate the forecasts; results in the second and third rows are obtained using the trivariate models that add repo rates and Treasury yields to model (6). All models include 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 where the curve flattens around its upper asymptote (i.e., imposing the constraint: $q_{it} - q_t^* > \theta_1 - \theta_2/\sqrt{3}$). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix D.