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

Working Paper

Scarce, abundant, or ample? A time-varying model of the reserve demand curve

Staff Report, No. 1019

Provided in Cooperation with:
Federal Reserve Bank of New York

Suggested Citation: Afonso, Gara; Giannone, Domenico; La Spada, Gabriele; Williams, John C. (2022) : Scarce, abundant, or ample? A time-varying model of the reserve demand curve, Staff Report, No. 1019, Federal Reserve Bank of New York, New York, NY

This Version is available at:
https://hdl.handle.net/10419/266103

STAFF
REPORTS

NO. 1019
MAY 2022

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

Abstract

Does the federal funds rate respond to shocks when aggregate reserves are in the trillions of dollars? Has banks' demand for reserves moved over time? We provide a structural time-varying estimate of the slope of the reserve demand curve over 2010-21. We estimate a time-varying vector autoregressive model at daily frequency with an instrumental variable approach to address endogeneity. Consistent with economic theory, our estimates show a nonlinear demand function that exhibits a negative slope in 2010-11 and 2018-19 but is flat over 2012-17 and after mid-2020. We also find that the curve has moved outward, both vertically and horizontally.

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, Thomas Eisenbach, Huberto Ennis (discussant), Antoine Martin, Andrea Tambalotti, and participants at the 2021 Banque de France Conference on Real-Time Data Analysis, Methods and Applications, and at the 2021 ECB Conference on Money Markets for valuable feedback. They also thank Valerie Baldinger, 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

Over the past fifteen years, reserves in the banking system have grown from tens of billions of dollars to several trillion dollars. This extraordinary rise in the supply of reserves poses a natural question: are the rates paid in the market for reserves sensitive to shocks when aggregate reserve holdings are so large? To address this question, this paper provides a structural time-varying estimate of the slope of the reserve demand curve at daily frequency over 2010-2021, when reserves ranged from $1 trillion to $4 trillion.

The reserve demand curve describes the price at which banks are willing to borrow and lend their reserve balances as a function of aggregate reserves in the system. The interest rate at which reserves are borrowed and lent is the federal funds rate, which is also the policy rate targeted by the Federal Open Market Committee (FOMC). The reserve demand function measures banks' demand for liquidity. Estimating the sensitivity of the federal funds rate to shocks to the level of reserves is of paramount importance for the implementation of monetary policy.

Estimating the price sensitivity of the demand for reserve is challenging for three reasons. First, theory predicts that the reserve demand curve is a highly nonlinear function. The demand curve can be divided into three regions (Poole (1968); Ennis and Keister (2008); Afonso et al. (2019)). When reserves are low and close to the aggregate requirements, the demand curve has a steep negative slope reflecting their scarcity value. When aggregate reserves are abundant, the reserve demand curve becomes flat around the interest on reserve balances (IORB) paid by the Federal Reserve. Between these two regions, market frictions generate a smooth transition – an intermediate regime of "ample" reserves where the demand curve exhibits a gentle downward slope.

Second, since the 2008 crisis, the demand for reserves has been affected by various structural changes in the regulation and supervision of banks, in banks' internal risk-management frameworks, and in the structure of the market for reserves itself. As a result, not only is the reserve demand curve highly nonlinear, but it may also have been subject to structural changes over the years.

Third, estimation of the demand for reserves is subject to potential endogeneity issues due to high-frequency confounding factors in the demand equation. Although the Federal Reserve no longer targets federal funds rates by adjusting aggregate reserves daily as it did before 2008, it may still respond to sudden dislocations in money-market rates by quickly changing the supply of 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 daily money market conditions, such as the balance of the U.S. Treasury's account with the Federal Reserve and usage of the Federal Reserve's overnight reverse repurchase agreement facility.

Our estimation strategy addresses all three of these challenges. Instead of estimating a highly nonlinear function with possible slow-moving structural shifts, we estimate the slope of a daily-frequency linear function with time-varying coefficients and stochastic volatility. This approach enables us to retrace the nonlinear shape of the curve over time by moving along the curve, while allowing for low-frequency movements of the curve, thus providing a flexible approach to addressing

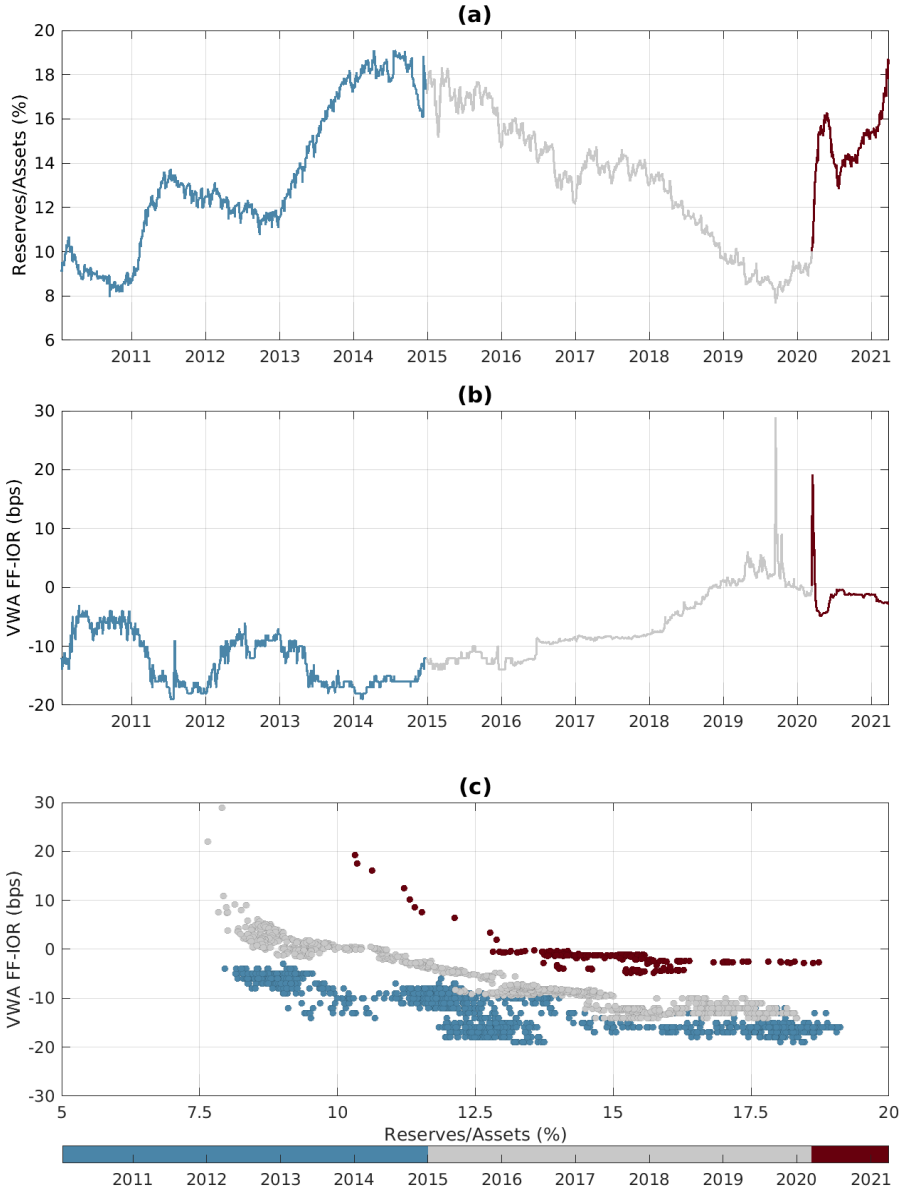


Figure 1: **Reserves, federal funds rates, and the reserve demand curve.** Panel (a) plots aggregate reserves relative to commercial banks' assets from January 1, 2010 to March 29, 2021. Panel (b) shows the spread between the volume-weighted average federal funds rates and the IORB rate (in basis points). Panel (c) plots the relationship between spreads 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 federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at the Federal Reserve Bank of New York. Weekly data on total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances is available from FRED ("IOER").

the first two challenges. Our specification is agnostic about the economic forces moving the curve over time, which allows for very general types of structural changes; the only important assumption is that the parameters of the time-varying linear model evolve more slowly than the daily liquidity shocks affecting banks' demand for reserves.

We address endogeneity by using daily data and an instrumental variable (IV) approach combined with a time-varying vector autoregressive (TV-VAR) model of the joint dynamics of reserves and federal funds rates. The use of daily data allows us to directly control for the window-dressing of European banks around month ends, which is an important but periodic and transient demand shock. To control for more general confounding factors, we instrument reserves in our linear time-varying demand specification with their forecast errors from the TV-VAR model. By estimating this model on daily data, we can control for the response of the Federal Reserve to past dislocations in federal funds rates. Moreover, to ensure that our instruments are not contaminated by transient confounding factors due to non-reserve Federal Reserve accounts, we use forecast errors from five days before. The instrument's relevance stems from the persistence of the reserve path over time.

Panel (a) of Figure 1 shows the time evolution of aggregate reserves normalized by banks' total assets to control for the growth of the banking industry in our sample. Reserves went through a full expansion-contraction cycle from 2010 to late 2019 and expanded again in early 2020, 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 to control for changes in the monetary policy stance, which mechanically shift the curve up and down by moving its lower bound. 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 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 clear nonlinear, downward-sloping relationship between prices and quantities that moves 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.

Of course, Figure 1 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. Consistent with the predictions of the theory and the evidence in Figure 1, our IV estimates show a nonlinear demand curve: the slope is significantly negative in 2010-2011 and 2018-2019 and zero in the interim and end periods. This time variation in the curve's slope reflects movements along the curve driven by small exogenous supply shocks captured by our instrument.

The curve itself, however, also moves horizontally at a low frequency. In the earlier part of our sample, the rate sensitivity to reserve shocks fades as reserves exceed 12% of banks' assets; in the latter part, instead, it reemerges around 13%, suggesting a moderate shift to the right. Below these thresholds, the curve's slope gets steeper as reserves decrease, consistent with the theory. A one-percentage-point drop in normalized reserves increases the federal funds-IORB spread by 1.3 basis points (bp) in 2010 and by 1.1 bp in 2019, while having no effect in 2012-2017 and after March

2020. These estimates are robust to controlling for spillovers from the repo and Treasury markets. Results are qualitatively similar if we normalize reserves by GDP instead of banks' assets.

Our structural time-varying estimates of the slope tell us where the reserve demand curve transitions from being flat (abundant reserves) to being gently sloped (ample reserves) and suggest a modest horizontal shift to the right over the last ten years. They do not, however, tell us much about possible vertical shifts, as vertical shifts do not affect the slope. To illustrate how the curve may have moved upward over the years, we use a post-processing methodology based on our forecasting model and the predictions of the theory. We fit a nonlinear demand function with horizontal and vertical shifts on repeated cross-sections of the joint forecasts of rates and reserves from the TV-VAR. Our results suggest that vertical upward shifts are present and more important than horizontal shifts in the last part of the sample. This observation implies that the level of the federal funds-IORB spread may not be a good summary statistic for the rate elasticity to reserve shocks. Our results also indicate that as reserves approach 9-11% of banks' assets, the rate elasticity reaches its maximum growth rate, suggesting a transition between ample and scarce reserves.

A key additional advantage of our methodology is that it provides a flexible framework that can be used to monitor the market for reserves in real time and to assess the relative scarcity or ampleness of the supply of reserves. We conduct an extensive analysis of the external validity of our forecasting multivariate model by estimating it recursively to mimic its use as a policy monitoring tool. We find that the out-of-sample predictions are similar to the in-sample predictions and that the model generates accurate forecasts based on several different metrics.

This paper adds to an extensive literature on the demand for reserves. Hamilton (1997) estimates the slope of the reserve demand curve with a time-invariant model and daily data over the relatively short period of 1989-1991. The econometric approach we develop in this paper enables us to estimate the curve and its evolution over a long period characterized by prevalent institutional changes.

The most recent literature has focused on reserve demand after the 2008 financial crisis. Smith (2019) estimates the slope of the demand curve over 2014-2018 using a time-invariant VAR model on weekly data. Lopez-Salido and Vissing-Jorgensen (2022) argue that bank deposits are a central driver of reserve demand. Using monthly data over 2009-2021, they estimate a time-invariant relationship between the federal funds-IORB spread and reserves adjusted for deposits; they use this estimated time-invariant demand function to discuss the runoff of the Federal Reserve balance sheet. Smith and Valcarcel (2022) estimate a time-varying VAR using weekly data over the 2017-2019 period to analyze the effect of the runoff of the Federal Reserve balance sheet on financial markets. Finally, the consequences of a central bank's balance-sheet expansions on market liquidity are studied in Acharya and Rajan (2022), which presents a theoretical framework where balance-sheet expansions need not necessarily enhance the net availability of liquidity in the banking system.

The rest of this paper is organized as follows. Section 2 discusses the theory and institutional setting. Section 3 describes the data. Section 4 describes the model, endogeneity problems, and identification strategy. Section 5 reports the results of our IV estimation. Section 6 reports the results of our post-processing fit of the whole demand curve. The Appendix includes a detailed description of the forecasting model, its out-of-sample validation, and several robustness checks.

2 The Demand for Reserves

2.1 Theory

Reserves are balances that depository institutions ("banks") maintain with the Federal Reserve to satisfy regulatory requirements and liquidity needs, such as daily payments to other institutions. Banks borrow and lend reserves in the federal funds market, typically overnight, at the federal funds rate. The reserve demand curve describes the price at which banks are willing to trade reserves as a function of the total amount of reserves in the banking system.

For each day t, we can write this curve as

$$p_t = p_t^* + f(q_t - q_t^*; \theta) + \varepsilon_t, \tag{1}$$

where p is the federal funds rate (price), q is the aggregate reserves (quantity), and ε is a daily demand shock. $f(\cdot;\theta)$ is a continuous weakly decreasing nonlinear function parametrized by θ, with both an upper and a lower bound. For normalization, we set $\lim_{x\to\infty} f(x;\theta) = 0$, so that p^* is the lower bound of the demand curve as reserves go to infinity. The variable p^* pins down the vertical location of the curve, which we allow to slowly move over time (i.e., at lower frequencies than daily). The variable q^* represents the slow-moving horizontal location of the curve. We assume that θ (i.e., f's functional form) is fixed over time.

Where does equation (1) come from? The economic theory behind the reserve demand curve can be summarized as follows. Assume that banks face some uncertainty about their end-of-day balances and trade in the federal funds market during the day to insure themselves against this uncertainty. For very high levels of reserves, banks are not concerned about breaking their liquidity requirements, and the demand curve is flat around the interest rate paid by the Federal Reserve on reserve balances (the IORB rate), which is the banks' opportunity cost of lending in the federal funds market. No bank, in fact, should be willing to lend at a lower rate and competition among lenders should drive federal funds rates to the IORB rate. As a result, when reserves are abundant, even sizable shocks to their aggregate level have no effect on rates.

As reserves drop below some threshold, banks start to be concerned about meeting their regulatory requirements and liquidity needs; as a result, rates move away from their lower bound, and the demand curve begins to display a negative slope. As reserves keep decreasing, the slope of the curve, i.e., the sensitivity of rates to reserve shocks, becomes increasingly steeper reflecting the increasing scarcity value of reserves.

Below some threshold, however, the curve must flatten, and its slope decrease. The reason is that, for extremely low reserve levels some banks must borrow from the Federal Reserve's Discount Window (DW), and the federal funds rate must converge to the DW rate. No bank, in fact, should be willing to borrow at a higher rate, and competition among borrowers should drive rates to the DW rate.

In this paper, we are interested in the right part of the curve, where reserves transition from being abundant to scarce. The reason is that the federal funds rate was significantly below the DW

rate throughout our sample, suggesting that the market operated away from the extreme scarcity region around the curve's upper asymptote.[1] Figure 2 shows the part of the reserve demand curve we focus on. The intermediate region between scarcity and abundance is referred to as the region of ample reserves, and identifying its possibly time-varying range is one the main goals of this paper.

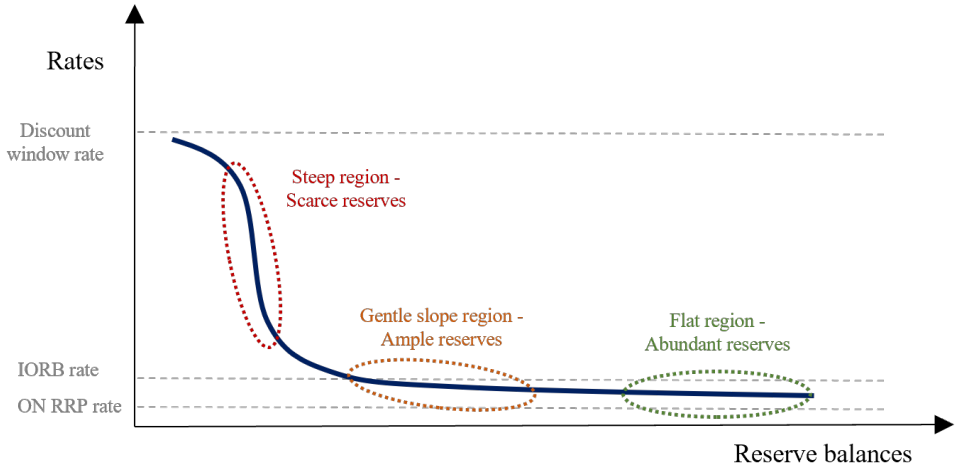


Figure 2: **Reserve demand curve.**

2.2 Monetary policy implementation

Whether reserves are abundant, ample, or scarce has important implications for monetary policy implementation because the FOMC uses the federal funds rate to communicate the monetary policy stance. Indeed, these regions map into the Federal Reserve's monetary policy frameworks before and after the 2008 crisis.

Before the global financial crisis, total reserves were in the tens of billions and did not earn interest (i.e., the IORB rate was set at zero). In this scarce reserves environment, the demand curve exhibited a steep downward slope, and the Federal Reserve kept rates close to the FOMC target by adjusting the supply of reserves through relatively small daily open market operations. The Federal Reserve can change the level of aggregate reserves by trading securities with banks. When the Federal Reserve purchases securities from banks, adding assets to its balance sheet, it issues reserves by crediting the Federal Reserve accounts of the selling banks. Vice versa, when the Federal Reserve sells securities to banks, aggregate reserves decline.

To support the economy in the aftermath of the crisis, the Federal Reserve purchased large quantities of long-term assets from banks between 2008 and 2014. As a result, reserves exceeded a trillion in 2009 and peaked at $2.8 trillion in 2014. With abundant reserves, small adjustments to

[1]The only exception occurred on March 16, 2020, when the DW rate was decreased by 50 bp relative to the IORB rate to encourage DW borrowing amidst the money-market turmoil triggered by the Covid crisis. As a result, the federal funds rate equaled the DW rate for three days but returned to be significantly lower right afterwards.

the reserve supply do not influence rates because the demand curve is flat around the IORB rate. Fpr this reason, since 2008, the Federal Reserve has changed the IORB rate to affect rates and implement monetary policy (Keister et al. (2008)).[2]

Changes in the IORB rate, by changing the banks' opportunity cost of lending in the federal funds market, can be thought of as vertical shifts in the reserve demand curve. To control for these demand shifters, which are unrelated to structural changes in banks' demand for reserves but simply reflect changes in the monetary policy stance, we express the price p in the demand curve (1) as the spread between federal funds rates and the IORB rate.

Although changes in the IORB rate have effectively influenced federal funds rates, the IORB rate is not a hard minimum at which all institutions are willing to lend. Federal Home Loan Banks (FHLBs), which represent more than 85% of total lending in the market, are ineligible to earn interest on their Federal Reserve balances and are therefore willing to lend at lower rates. To set a hard floor, in 2013, the Federal Reserve introduced the ONRRP facility; through this facility, FHLBs and money market funds (MMFs), which lend to banks in other key money markets, can invest at the Federal Reserve via overnight repos at a fixed-rate below the IORB rate. The FOMC adjusts the ONRRP and IORB rates together in the implementation of monetary policy.

The segmentation of the federal funds market between banks and FHLBs suggests that, under our price normalization, the lower bound of the demand function (1), p^*, should range between the ONRRP-IORB spread and zero. Of course, other institutional frictions and structural changes can move p^* (and q^*) over time, and we provide some examples below.

2.3 Structural changes after 2008

Several structural changes have affected banks' demand for reserves since the 2008 crisis. The post-crisis regulatory framework is an important example. The Liquidity Coverage Ratio (LCR), introduced in the U.S. in 2015, requires that banks hold enough high-quality liquid assets, such as reserves or Treasuries, to survive a significant stress scenario lasting for one month. While the LCR does not explicitly impose that banks hold reserves, Ihrig et al. (2020) reports that large banks tend to use reserves to satisfy the LCR requirement.

Supervisory liquidity stress tests have also become prime drivers of banks' demand for reserves since 2008. Under Regulation YY's enhanced prudential standards introduced in 2014, large financial firms are required to hold liquidity buffers to cover outflows on the first day of the stress-test scenario, without reliance on the Federal Reserve. The demand for reserves has also increased due to the introduction of Dodd-Frank regulation that requires large firms to prepare plans ("living wills") describing their potential orderly resolution.

An additional driver of reserve demand is the lack of depth in late-day funding markets. After reserves became abundant, trading incentives declined, and average daily volume dropped from a pre-crisis level of $220 billion to $70 billion. Since then, bank-to-bank intraday lending, which

[2]The Federal Reserve began to pay interest on depository institutions' reserve balances in October 2008.

was used to insure against late liquidity shocks, has dried up, likely pushing banks' precautionary demand for reserves outward. Indeed, survey data collected by the Federal Reserve show that, in recent years, banks identify meeting large intraday payments as a major driver of their demand.[3]

These structural changes could be interpreted as the horizontal shifts q^* in the demand curve (1): for every level of the federal funds rate, they imply an increase in the quantity of aggregate reserves demanded by the banking system. As a result of these shifts, the level of reserves at which the curve stops being flat and start displaying a negative slope may have moved over time.

Structural changes, of course, may have also moved the reserve demand curve vertically by altering the opportunity costs and bargaining powers of market participants. Post-crisis regulations that impose a cost on the size of banks' balance sheets reduce the price at which banks are willing to borrow reserves for any level of aggregate reserves, pushing p^* down; examples of such regulations are the Basel III leverage ratio and the Federal Deposit Insurance Corporation's (FDIC) assessment fee (Kim et al. (2020)). The introduction of the ONRRP facility, instead, offers FHLBs and MMFs the opportunity to invest with the Federal Reserve through overnight repos, likely increasing their bargaining power over borrowing banks and leading to higher federal funds rates for any reserve level (Schulhofer-Wohl and Clouse (2018)). For this reason, persistent increases in the ONRRP rate relative to the IORB rate, such as those occurred over June 2018-May 2019, would improve FHLBs' and MMFs' bargaining power over time, pushing p^* up.[4]

2.4 Endogeneity

To understand the sources of endogeneity in the demand for reserves, it is important to understand how aggregate reserves can change. The level of reserves in the banking system changes for two reasons: either because the Federal Reserve buys or sells assets from banks, which changes the size of the Federal Reserve balance sheet, or because funds are transferred between reserves and non-reserve accounts at the Federal Reserve, which changes the composition of the Federal Reserve's liabilities keeping the size of the balance sheet constant. Some institutions have non-reserve Federal Reserve accounts and transact with private banks; when these transfers occur, reserves are either created or destroyed outside the Federal Reserve's control.[5] Both types of reserve fluctuations are associated with endogeneity problems.

The first type of endogeneity is due to the Federal Reserve's actions. The Federal Reserve responds to volatility in the federal funds market by adjusting the reserve supply to keep the federal funds rate within its target range, as it happened in September 2019 and March 2020.[6] These

[3]Summaries of the surveys' findings can be found at: `https://www.federalreserve.gov/data/sfos/sfos-release-dates.htm`. Since 2018, the Federal Reserve has conducted seven Senior Financial Officer Surveys (SFOS) to gather information on banks' reserve management strategies and practices.

[4]During this time period, the IORB-ONRRP spread was reduced three times, by 5 bp each time.

[5]In contrast, federal funds transactions between banks do not affect aggregate reserves; they just redistribute them.

[6]Before 2008, the endogeneity due to the Federal Reserve's actions was more severe because the Federal Reserve regularly adjusted the reserve supply through daily open market operations to control the federal funds rate. This endogeneity has disappeared since the Federal Reserve started using administered rates to implement monetary policy.

responses are quick and are put in place within a matter of days. In September 2019, for example, the federal funds rate spiked up on the 16^{th} and 17^{th}; starting on the 18^{th}, the Federal Reserve expanded the reserve supply by providing cash in the repo market, and rates returned to their prior levels (Afonso et al. (2020a)).

The second type of endogeneity is due to the activity of non-reserve Federal Reserve accounts that are correlated with money-market conditions and the demand for reserves. An example is the Treasury General Account (TGA) with the Federal Reserve. When banks buy Treasuries at a Treasury auction, they transfer funds from their Federal Reserve accounts to the TGA, which results in an increase in the TGA balance and a decrease in reserves. Around the same time, their demand for short-term funding increases as they finance their Treasury purchases with overnight repos. The temporary increase in repo rates can push the demand for reserves up because overnight federal funds and repos are close substitutes; in other words, federal funds rates increase not only because reserves decline but also because repo rates increase (Schulhofer-Wohl and Clouse (2018)).[7]

Martin et al. (2019) point out that the issuance of Treasuries also affects the federal funds rate by placing upward pressure on Treasury yields. The reason is that when Treasury yields go up, 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 wholesale deposit rates, pushing the federal funds rate up.

Other examples of non-reserve Federal Reserve accounts correlated with money-market conditions are the ONRRP facility and FHLB balances. When a MMF invests overnight in the ONRRP facility, it instructs its custodian bank to make a transfer to the ONRRP account; the result is a decrease in the bank's reserves and an increase in the ONRRP balance. Similarly, FHLBs do not hold reserves with the Federal Reserve, which is the reason why they do not earn the IORB rate on their balances, but use their Federal Reserve accounts to lend to banks in the federal funds market; when they do so, their Federal Reserve balances decrease, and aggregate reserves increase.

One way variations in the ONRRP and FHLB balances 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 market and from MMFs (Banegas and Tase (2020)). This window dressing has two effects on the market for reserves. First, it lowers rates as the demand for borrowing declines. Second, it lowers aggregate reserves because FHLBs lend less in the federal funds market, and MMFs compensate the reduction in European banks' demand for funding by investing more in the ONRRP facility.

More generally, changing conditions in the repo and Treasury markets, in addition to affecting banks' demand for reserves, also affect the level of reserves by changing the incentives of FHLBs to

[7]Another confounding factor related to the TGA are tax payments. When MMF investors use their shares to pay taxes, the MMF 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, MMFs also tend to reduce their overnight repo lending, which can lead to a temporary rise in repo and federal funds rates. Both Treasury settlements and tax payments played important roles in the money-market turmoil of September 2019 (Afonso et al. (2020a)).

lend in the federal funds market and of MMFs to invest in the ONRRP facility. These institutions, in fact, hold a sizable share of their portfolios in Treasuries and are the main lenders to banks in the overnight repo market. Overnight Treasury repos are especially relevant because they are a close substitute to overnight federal funds borrowing, and their relative importance in both banks' funding demand and MMF's lending supply has increased significantly after the 2014 SEC reform of the MMF industry (Cipriani and La Spada (2021); Anderson et al. (2020)).[8]

The endogeneity due to non-reserve Federal Reserve accounts has become more severe since 2008. One reason is that the repo and Treasury markets have increased substantially over the past decade. For example, in the Tri-party segment of the repo market, the volume of lending collateralized by Treasuries and agency debt increased from $1.3 trillion in May 2010 to $1.9 trillion from in March 2021, reaching a peak of $2.4 trillion in March 2020.[9] Another reason is that the variability of the non-reserve accounts has also increased significantly since the 2008 crisis (Afonso et al. (2020b)). Figure 3 shows that, excluding currency in circulation, non-reserve liabilities amounted to 10% of aggregate reserves in November 2013 and to 90% in July 2020. The TGA increased from $90 billion at the end of 2010 to $350 billion at the end of 2019, a hike from 9% to 20% of aggregate reserves. In July 2020, amidst the response to the Covid crisis, the TGA jumped to $1.8 trillion, reaching almost 70% of total reserves. The ONRRP facility only opened in September 2013; since then, its balance has fluctuated between less than a billion and more than $400 billions, reaching 20% of aggregate reserves in late December 2016.

In Section 4, we propose a general instrumental-variable methodology to control for the endogeneity issues caused both by the activity of non-reserve Federal Reserve accounts and by the Federal Reserve's supply of reserves.

3 Data

We use daily data from January 1, 2009 to March 29, 2021. Since the federal funds market is only open on business days, we drop weekends and all federal holidays, including Mondays following holidays that fall on a Sunday.[10]

Using daily data is key for our identification strategy because it allows us to directly address the endogeneity issues due to the "window dressing" by 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 equation, we exclude one-day windows around month-ends.

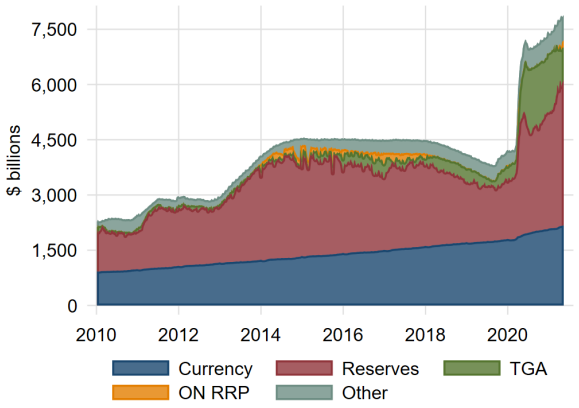
Our main variables of interest are aggregate reserves and federal funds rates. To calculate

[8]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.

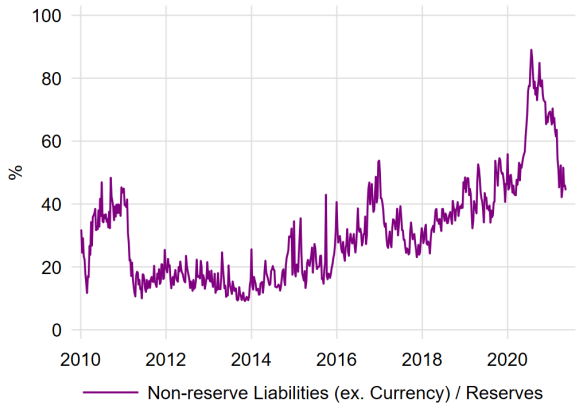
[9]These data are from the Data and Statistics Group at the Federal Reserve Bank of New York and; see `https://www.newyorkfed.org/data-and-statistics/data-visualization/tri-party-repo#interactive/volume`.

[10]We use the standard 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`.

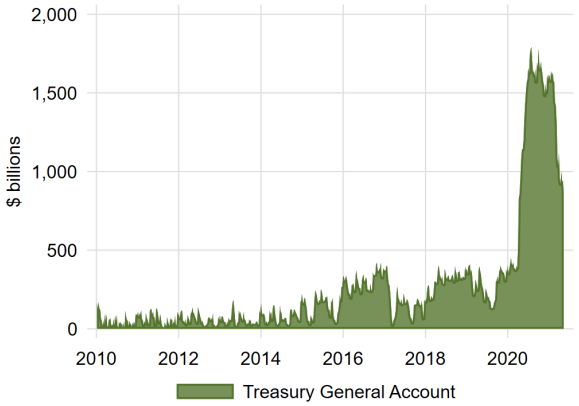
(a) Federal Reserve's liabilities



(b) Ratio of non-reserve liab. (exc. currency) to reserves



(c) TGA



(d) ONRRP

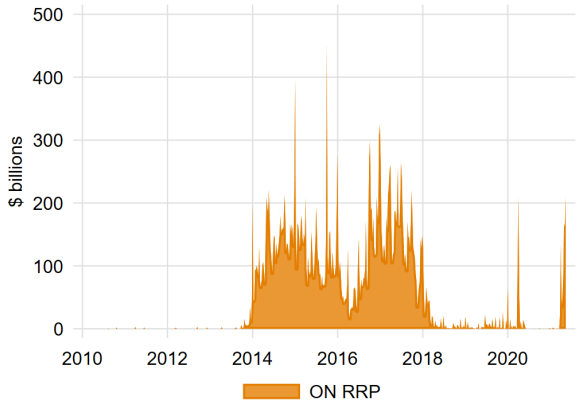


Figure 3: **Federal Reserve Liabilities.** Evolution of the Federal Reserve's liabilities (panel (a)), ratio of non-reserve liabilities (excluding currency in circulation) to aggregate reserves (panel (b)), Treasury General Account (TGA) (panel (c)), and Overnight Reverse Repurchase Agreement (ONRRP) facility (panel (d)) from January 2010 to March 2021. Data are weekly from Federal Reserve Statistical Release H.4.1.

aggregate reserves, we use daily confidential data on aggregate reserve balances held by depository institutions; these data are provided by the Federal Reserve Bank of New York.[11] To take into account the significant growth of the banking system over our sample, we normalize reserves by commercial banks' assets. Data on banks' assets are publicly available at weekly frequency from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). We linearly interpolate weekly data to obtain a daily series of commercial banks' assets.

For the federal funds rate, we use the daily volume-weighted average based on transactions data, published by the Federal Reserve Bank of New York. The underlying transactions data are the same as used for the official calculation of the effective federal funds rate (EFFR). As explained in Section 2.2, to control for changes in the monetary policy stance, we subtract the IORB rate from the volume-weighted average federal funds rate. We use daily interest rate on excess reserves, publicly available from FRED ("IOER").

In some of our 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).[12] 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 Treasury securities at one-year constant maturity; data are publicly available from FRED ("DGS1"). Finally, data on the daily rate at the Federal Reserve's primary credit DW program are also publicly available from FRED ("DPCREDIT").

4 Empirical Implementation

4.1 Empirical model and estimation

As discussed above, estimating the sensitivity of federal funds rates to reserve shocks is challenging for three reasons: (i) the reserve demand curve is a nonlinear function of aggregate reserves, which means that the slope is itself a function of aggregate reserves; (ii) after the 2008 crisis, persistent structural changes may have moved the curve over time; and (iii) potential endogeneity issues due to the Federal Reserve's response to dislocations in the federal funds market and due to high-frequency confounding factors.

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:

$$p_t = \alpha_t + \beta_t q_t + \sigma_t v_t, \tag{2}$$

where p and q are the price and quantity of reserves, and v is a daily demand shock. To account

[11] A weekly version of these data is available in the H.4.1 report ("Reserve balances with Federal Reserve Banks").
[12] `https://www.dtcc.com/charts/dtcc-gcf-repo-index`.

for the growth of the banking sector over time, we measure q as aggregate reserves divided by commercial banks' total assets; to control for changes in the monetary policy stance, we measure p as the spread between the average federal funds rate and the IORB rate. All parameters in (2), including the variance σ of the demand shocks, are allowed to vary at daily frequency. The time-varying slope β_t measures the elasticity of rates to reserves on each day.

Model (2) is a locally linear approximation of the reserve demand curve (1) implied by the theory. Every day, we estimate a straight line; as time passes, these lines move and retrace the demand curve. This approach enables us to capture the nonlinear nature of the curve, without specifying a particular functional form and allowing for low-frequency structural shifts.

Changes in the parameters of model (2) 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 (2) is that its parameters evolve more slowly than the liquidity demand shocks that hit banks on a daily basis, which allows us to disentangle (low-frequency) variation in β from (high-frequency) variation in v. This assumption is plausible for two reasons. First, it took banks months to adjust to the post-crisis changes in regulation, supervision, and functioning of the reserves market described in Section 2. 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 expansion-contraction path 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 (2).[13] 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):

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

where q and p are defined as in equation (2), 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 Ω_t, i.e., $(u_{q,t}, u_{p,t})' \sim \mathcal{N}(0, \Omega_t)$.

We estimate model (3) at the daily frequency using Bayesian methods; each parameter is modeled as a stochastic process. Consistent with our approximate reserve demand curve (2), the basic assumption behind the estimation of model (3) is that its parameters evolve more slowly than the daily errors. Specifically, as Primiceri (2005), we assume that the parameters follow slow-moving

[13] 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. More recently, Del Negro et al. (2020) use a similar approach for the estimation of the Phillips curve.

random walks, whose innovations are uncorrelated with the u errors at all leads and lags. Given the data, our estimation gives us the joint posterior distribution of the parameters (c's, b's, and Ω) 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 (2), we can write the IV estimate of β_t as

$$\beta_t^{IV} = \frac{\text{cov}(p_t, u_{q,t-h})}{\text{cov}(q_t, u_{q,t-h})}. \tag{4}$$

Our estimation of the forecasting model (3) gives us simultaneously the forecast errors u and their time-varying covariances with the observables 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 A.1, we describe in detail how to conduct inference on β_t^{IV} and explain how the covariances in equation (4) 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.

Finally, inference on β_t^{IV} depends on the choice of the forecast horizon h. There is a clear trade-off between the exogeneity of the instrument and the precision of the estimate. The longer is the horizon, the more plausible is the exogeneity assumption. A longer horizon, however, also 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 exogeneity assumption underlying our identification strategy is that the forecast error for reserves at time t from model (3) is uncorrelated with the shock at time $t + 5$ in the demand equation (2). 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. A sufficient condition for this exclusion restriction is the absence of contemporaneous correlation between reserve supply shocks and daily demand shocks. This condition, however, is not necessary; 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 of our instrument is plausible for two reasons.

First, since 2008, the Federal Reserve has changed the IORB rate to implement monetary policy, rather than adjusting the reserve supply via daily open market operations. In this monetary policy framework, the Federal Reserve's supply does not reflect a contemporaneous reaction to changing conditions in the federal funds market; it only responds to past price dislocations, quickly but with a delay of at least a day (Afonso et al. (2020a)). For this reason, it is important to use daily data to extract exogenous variation in the reserve supply. By using average data at lower frequencies, say weekly or monthly, variation in reserves would mix supply and demand shocks.

In principle, the demand shocks in equation (2) could be serially correlated; in this case, today's change in the Federal Reserve's supply, which is a response to yesterday's demand shock, could be correlated with today's demand shock. By conditioning on past daily data up to ten lags and allowing for time-varying coefficients, our forecasting model can control for the Federal Reserve's response function to price volatility; as a result, its errors should be residualized with respect to this confounding factor even if demand shocks were correlated. Moreover, even if our forecasting model were misspecified, and today's error for reserves were contaminated by a contemporaneous demand shock, the error from five days before should not; the reason is that, in our sample, the Federal Reserve's actions have effectively stabilized price dislocations quickly, typically within days.

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

As we explain in Section 2, conditions in the repo and Treasury markets are particularly important confounding factors. To further strengthen our identification, in robustness checks, we 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 the federal funds market. The estimation of these trivariate VAR models is similar to that of the baseline bivariate model; details can be found in Appendix A.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 total supply of reserves and then expand 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 A.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 (4)) 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 distributions of β_t^{IV} already reflects the uncertainty in both the numerator (i.e., the reduced-form coefficient) and the denominator (i.e., the first-stage coefficient). In this way, our inference directly accounts for the possible non-normality of the IV estimate in equation (4), which could occur if the denominator is close to zero.

Our IV estimation is also robust to autocorrelation and heteroschedasticity of the demand shocks in equation (2). The reason is that the elasticity β_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 (3) is very flexible and general because the parameters of the model, as well as the variances and covariances of the residuals, are allowed to change over time. Thanks to this feature, the model can account for complex and evolving dynamic interactions between the federal funds rate and reserves. In many contexts, however, high model complexity comes with the curse of dimensionality; that is, the model tends to overfit in-sample and perform poorly out-of-sample. Several papers have shown that, when used for macroeconomic forecasting, the class of models to which model (3) belongs does not suffer from this pathology and can produce accurate forecasts even in real time with quarterly macroeconomic data (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 federal funds rates and reserves five days ahead; it shows that model (3) is not affected by the curse of dimensionality as the in-sample and the out-of-sample predictions are similar and close to the realized data. In Appendix A.4, we formally evaluate the out-of-sample accuracy of the model's predictions of the model in real time under various metrics (root-mean-square error, log-scores, probability integral transform). We find that the out-of-sample predictions are similar to the in-sample predictions and show that the model provides reliable and stable inference. The 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.

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 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 March 2021 (expansion). Consistent with economic theory, the relationship between the federal funds rate and aggregate reserves is nonlinear in all periods: it is negative for levels of reserves that are sufficiently low, say below 10-11% of banks' assets, whereas it is flat if reserves are sufficiently large, say above 14-15% of banks' assets.

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-

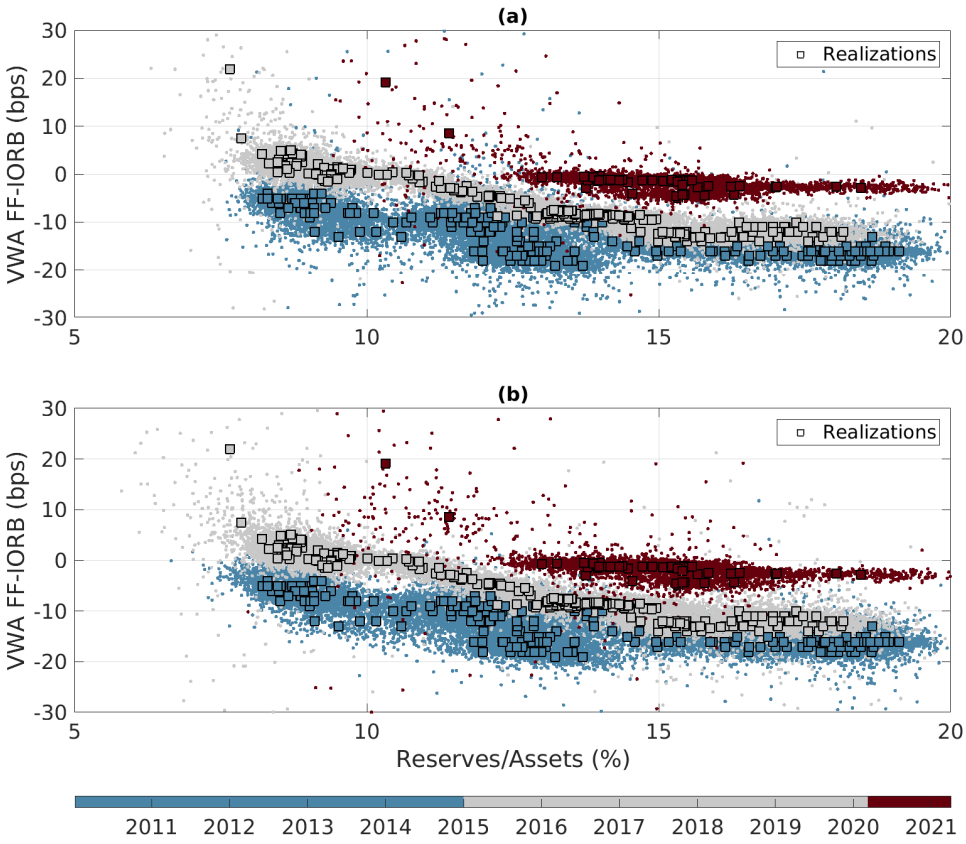


Figure 4: **In-sample (a) and out-of-sample (b) joint forecasts of federal funds rates and reserves five days ahead.** The forecasts are generated using the bivariate time-varying model (3), drawing 100 times every five business days from the joint posterior distribution of reserves and rates for each 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. Federal funds rates are 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 federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at the Federal Reserve Bank of New York. Weekly data on total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances is available from FRED ("IOER").

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. This shift is consistent with the regulatory and supervisory changes developed in response to the 2008 crisis and implemented over the following years. As discussed in Section 2, these changes may have increased banks' demand for reserves at every price level. Since this increase is due to permanent structural shocks, rather than to high-frequency liquidity shocks, it is captured by the slow-moving time-varying coefficients of our forecasting model.

More importantly, over 2010-2020, the relationship between the federal funds rate and aggregate reserves seems to mainly move vertically. As shown in Figure 4, the 2015-2020 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. This upward shift is consistent with an increase in the bargaining power of FHLBs and MMFs, key lenders to banks in the overnight funding markets, due to the introduction of the ONRRP in late 2013, among other factors.

A second and even more relevant vertical shift is visible for the last period. From March 2020 to March 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. This sizable upward shift is consistent with a steady increase of FHLBs' and MMFs' bargaining power over time, as well as with an increase in market risk aversion and persistent uncertainty caused by the Covid-19 crisis.

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; in Section 6, we quantify the horizontal and vertical shifts through a post-processing methodology and discuss in greater detail their possible economic interpretations.

5.2 The slope of the reserve demand curve and the region of ample reserves

5.2.1 OLS estimation

We now turn to estimating the time-varying slope of the reserve demand curve, i.e., the elasticity β_t in equation (2). 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 (3) 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 aggregate reserves: it is negative up to mid-2014 as reserves went from \$1 to \$2.7 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 goes back to zero after March 2020 as the Federal Reserve expanded the reserves supply above $3 trillion.

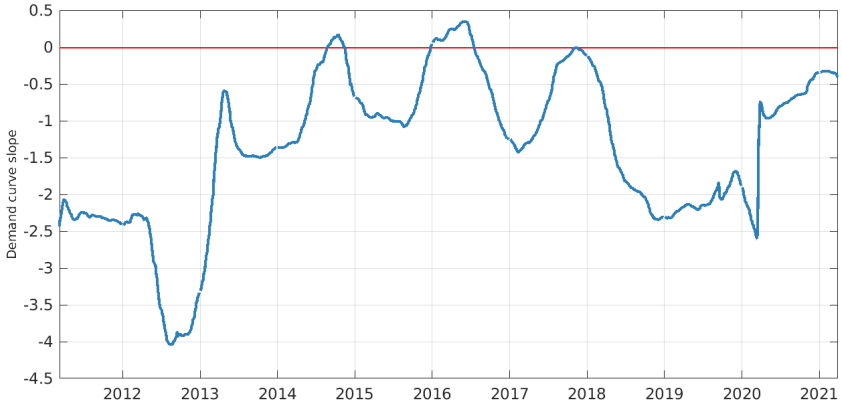


Figure 5: **OLS estimate of the elasticity of federal funds rates to reserves.** The slope of the reserve demand curve is estimated by running OLS regressions on rolling windows (244 business days) of in-sample forecasts of the spread between the federal funds rate and the IORB rate against in-sample forecasts of reserves normalized by banks' assets. Each time series excludes one-day windows around month-ends to control for the transient changes in the level of reserves and federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at the Federal Reserve Bank of New York. Weekly data on total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances is available from FRED ("IOER").

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 (4), together with their 95% and 68% confidence bands; panel (c) shows the same information for the IV estimate itself.

The numerator in equation (4), 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.

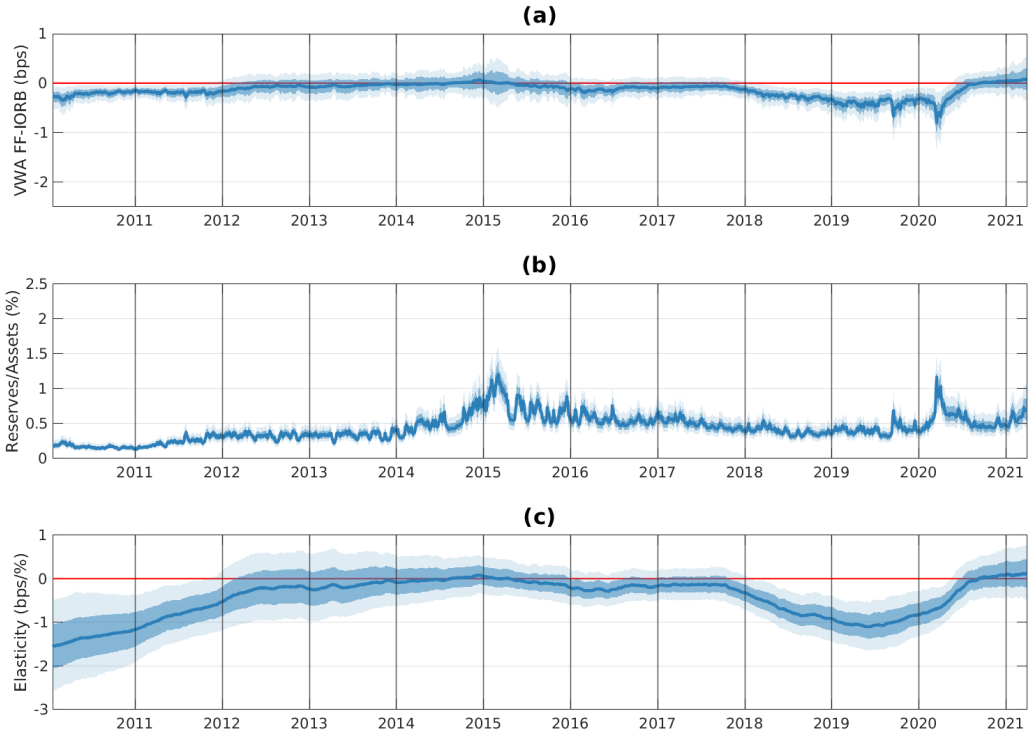


Figure 6: **In-sample IV estimate of the elasticity of federal funds rates to reserves.** The in-sample IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of federal funds rates (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (4). Forecast errors and impulse responses are estimated in-sample from model (3) with ten lags ($m = 10$ days). The solid blue line represents the posterior median. The dark and light blue shaded areas correspond to 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to commercial banks' assets. Federal funds rates are 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 federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at the Federal Reserve Bank of New York. Weekly data on total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances is available from FRED ("IOER").

The denominator in equation (4) 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 the ratio of (a) to (b), i.e., our IV estimate (4) of the sensitivity of the federal funds rate to reserve shocks from 2010 to 2021. Although different in magnitude, results 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 of rates and reserves. The rate elasticity was significantly negative but steadily increasing from 2010, when reserves ranged between 8% and 10% of banks' assets, to 2011, when reserves exceeded 12% of banks' 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 significant negative slope emerged again, as the Federal Reserve balance-sheet normalization led reserves to drop below 13% of banks' 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 reserves jumped above 16% of banks' assets, 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, β_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.

Taken together, the results in panel (c) of Figure 6 and in panel (a) in Table 1 are consistent with the nonlinear reserve demand curve predicted by the theory in Section 2. Our time-varying estimates of the elasticity of the federal funds rate to reserve shocks suggest that the slope of the reserve demand curve is itself a function of aggregate 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 GDP instead of banks' total assets, confirming that our choice of the normalization

2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
(a) Bi-variate Model											
-1.34	-0.82	-0.27	-0.14	-0.01	-0.06	-0.21	-0.16	-0.68	-1	-0.26	0.10
(-2.29,-0.36)	(-1.59,-0.09)	(-1.01,0.55)	(-0.83,0.6)	(-0.51,0.51)	(-0.47,0.38)	(-0.59,0.18)	(-0.54,0.23)	(-1.26,-0.13)	(-1.56,-0.43)	(-1.08,0.45)	(-0.44,0.72)
(b) Tri-variate Model with Repo Rates											
-1.97	-1.28	-0.85	-0.68	-0.13	-0.12	-0.22	-0.17	-0.63	-1.03	-0.33	-0.07
(-2.89,-1.02)	(-2.2,-0.34)	(-1.67,0.05)	(-1.57,0.16)	(-0.72,0.41)	(-0.57,0.35)	(-0.66,0.22)	(-0.61,0.28)	(-1.22,-0.07)	(-1.67,-0.37)	(-1.16,0.43)	(-0.72,0.66)
(c) Tri-variate Model with One-Year Treasury Rates											
-1.35	-0.91	-0.33	-0.22	-0.05	-0.06	-0.24	-0.12	-0.7	-1	-0.24	0.07
(-2.36,-0.34)	(-1.73,-0.07)	(-1.17,0.65)	(-0.99,0.64)	(-0.62,0.53)	(-0.52,0.42)	(-0.68,0.21)	(-0.55,0.34)	(-1.34,-0.04)	(-1.6,-0.37)	(-1.09,0.53)	(-0.5,0.77)

Table 1: **In-sample IV estimate of elasticity of federal funds rates to reserves by year.** The estimate of elasticity is obtained as the posterior median of the ratio between the impulse response of federal funds rates and the impulse response of reserves to a forecast error in reserves at a five-day horizon; see equation (4). In panel (a), forecast errors and impulse responses are estimated in-sample from the time-varying bivariate model (3) with ten lags ($m = 10$ days). In panel (b), they are estimated in-sample from an augmented trivariate version of the model that also includes daily repo rates, with ten lags ($m = 10$ days). In panel (c), elasticities are estimated in-sample from an augmented trivariate version of the baseline model that includes yields on 1-year U.S. Treasury securities, with ten lags ($m = 10$ days). The reported elasticities are calculated by year. Reserves are measured as a ratio to commercial banks' assets. Federal funds rates are 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 federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at 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 total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances and daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER" and "DGS1" respectively).

factor does not drive our results but simply removes a time trend in nominal reserves.

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 β_t^{IV} seems to come from small exogenous supply shocks captured by our instrument.

Our results, however, also suggest the presence of modest low-frequency horizontal shifts. In fact, the level of reserves at which the demand curve transitions from being flat to being negatively sloped 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 $1.6 trillion at the end of 2011 and $2.2 trillion at the beginning 2018. This modest shift to the right of the demand curve is consistent with an increase in the demand for reserves due to the regulatory and supervisory liquidity requirements implemented after the 2008 crisis.

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. Moreover, as we discussed in Section 4.3, the forecasting model we use to construct the time-varying instrument for aggregate reserves displays good predictive accuracy out-of-sample, enabling us to monitor conditions in the reserve market in real time. Real-time estimates of the curve's slope from our model can be interpreted as an early-warning signal for the transition from abundant to ample reserves.

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 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 (3) 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 (3) by including daily repo rates (relative to the IORB rate). 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.[14]

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 (3) with the spread between daily market yields on 1-year U.S. Treasury securities and the IORB rate. Then, we use the reserve forecast error from five days before as the instrument for reserves in our IV estimation. 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.

6 Implications for the Reserve Demand Curve

Our estimation methodology is highly flexible and able to identify the time-varying sensitivity of the federal funds rate to reserve shocks, but it does not directly allow for the recovery of the full 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 discussion in Section 2. We assume that the shape of the demand function is time-invariant but allow for slow-moving structural shifts in the vertical and horizontal locations of the function. In addition to providing an empirical description of the demand for reserves, this exercise also facilitates the assessment of the relative scarcity of the reserve supply, as discussed below.

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 transition between abundant and ample reserves, they do not answer two important questions. The first one is whether the demand curve has moved vertically in our sample period; our IV estimates

[14]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.

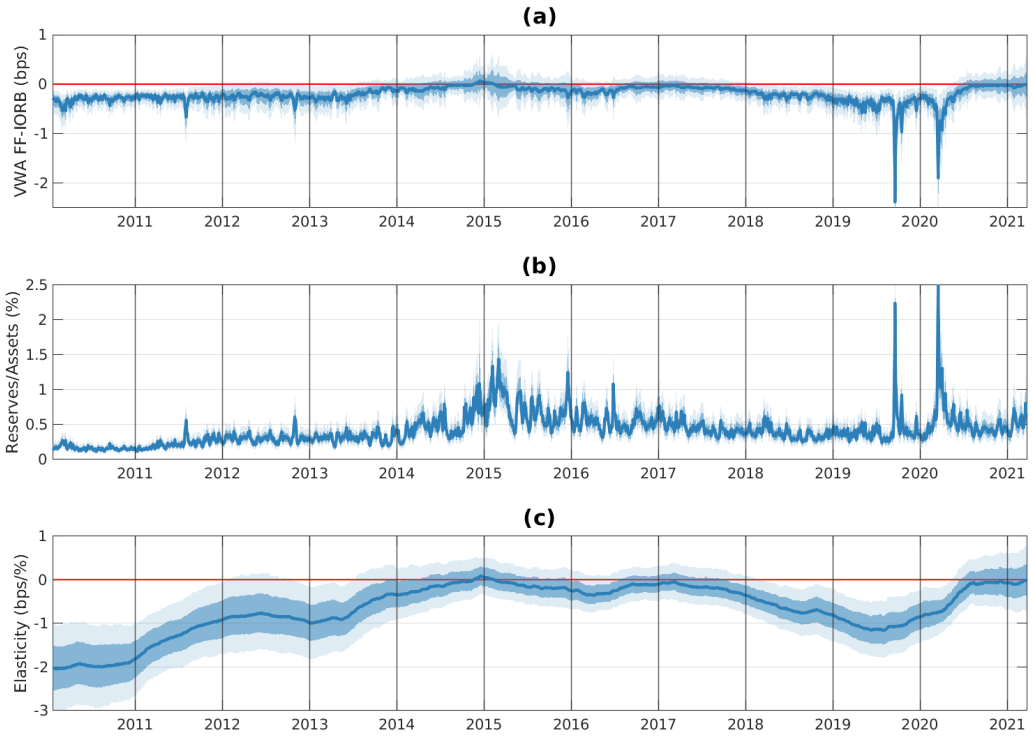


Figure 7: **In-sample IV estimate of the elasticity of federal funds rates to reserves controlling for repo rates.** The in-sample IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of federal funds rates (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (4). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (3) 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 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to commercial banks' assets. Federal funds rates are 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 federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at 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 total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances is available from FRED ("IOER").

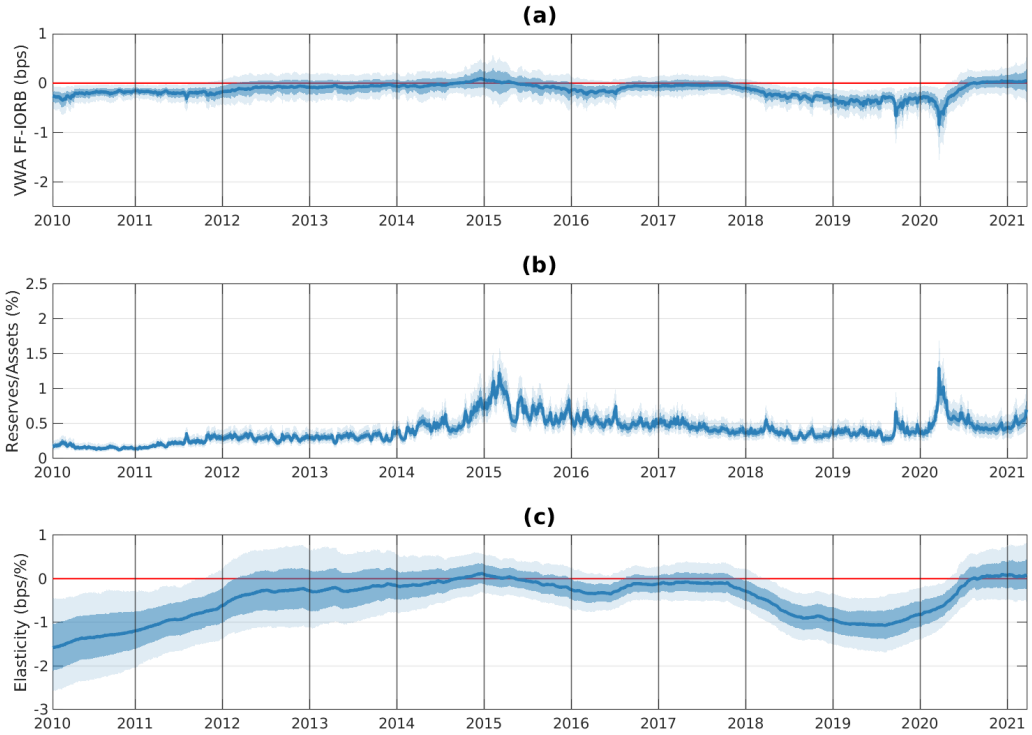


Figure 8: **In-sample IV estimate of the elasticity of federal funds rates to reserves controlling for 1-year U.S. Treasury securities yields.** The in-sample IV estimate of the elasticity (panel (c)) is obtained as the ratio between the impulse response of federal funds rates (panel (a)) and the impulse response of reserves (panel (b)) to a forecast error in reserves at a five-day horizon; see equation (4). Forecast errors and impulse responses are estimated in-sample using a trivariate version of model (3) 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 68% and 95% confidence bands. The elasticity is calculated daily. Reserves are measured as a ratio to commercial banks' assets. Federal funds rates are 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 federal funds rates caused by month-end window-dressing of European banks (see Section 2). Daily data on reserves and federal funds rates are provided by the Markets Group at the Federal Reserve Bank of New York. Weekly data on total assets for commercial banks in the U.S. and for U.S. branches and agencies of foreign banks are publicly available from the Federal Reserve Economic Data, FRED ("TLAACBW027SBOG"). Daily interest on reserve balances and daily yields on 1-year U.S. Treasury securities are available from FRED ("IOER" and "DGS1" respectively).

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 (2) either because the intercept of the linear approximation will change 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 rate level and the slope of the curve, which means that the federal funds rate cannot be used as a proxy for the rate elasticity to reserve shocks. After 2008, in fact, one may be tempted to use the rate level 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).[15] As a result, absent vertical shifts, an increase in the federal funds rate would imply an increase in the rate elasticity, suggesting that reserves are becoming scarcer. In the presence of vertical shifts, however, this one-to-one mapping no longer holds, and trends in the federal funds rate do not necessarily reflect changes in the rate elasticity.

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 reserve demand curve is flat; 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 (1), we specify the following demand function for reserves:

$$p_t = p_t^* + f(q_t - q_t^*; \theta) \qquad \text{with} \quad f(x;\theta) = \left(\arctan\left(\frac{\theta_1 - x}{\theta_2}\right) + \frac{\pi}{2}\right)\theta_3, \tag{5}$$

where p^* and q^* are the vertical and horizontal locations, θ_1 is a location parameter, θ_2 is a scale parameter, and θ_3 is a normalization factor. We choose this transformation of the arctan function because it has a smooth and decreasing sigmoid shape that goes to zero as $x \to \infty$, as predicted by the theory. 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 3/16/2020-3/29/2021. We assume that q^* and p^* do not change within each period; e.g., $q_t^* = q_1^*$ and $p_t^* = p_1^*$ for all t in 2010-2014.

[15] In the left part of the curve, instead, the absolute slope increases with reserves because the curve must flatten around the DW rate as reserves go to zero.

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

$$\sum_{k=1}^{3} \sum_{t \in T_k} \sum_{i=1}^{N} [p_{it} - p_k^* - f(q_{it} - q_k^*; \theta)]^2 \tag{6}$$

where T_1, T_2, and T_3 represent 2010-2014, 2015-3/13/2020, and 3/16/2020-3/29/2021; $i = 1, \ldots, N$ 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 (3).[16] 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 (6).

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 (5). Based on the economic theory and institutional framework discussed in Section 2, several factors can lead to structural shifts in the demand for reserves. In model (5), for example, p^* denotes the lower asymptote of the demand curve, which represents the wedge between the federal funds and IORB rates when reserves are abundant; factors affecting this wedge include market fragmentation, the introduction of the ONRRP facility, and bank regulations based on balance-sheet size.[17] Horizontal shifts in q^*, in contrast, reflect all those factors that shift out the demand curve at every price level, including changes in banks' liquidity-risk management such as the LCR and internal stress tests. 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 curve, θ_1 represents the point of maximum absolute slope, i.e., the level of normalized reserves at which the negative slope of the curve is the steepest. We can think of the region around θ_1 as the region of scarce reserves, where the federal funds rate is highly sensitive to even small reserve shocks. The point of maximum slope growth, instead, is $x = \theta_1 + \theta_2/\sqrt{3}$; this point is where the curve's absolute slope increases at the highest rate as reserves decrease, which we interpret as the threshold between ample and scarce reserves.[18]

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

[17] As discussed in Section 2, absent frictions, p^* should be zero.

[18] Since the arctan function is strictly decreasing everywhere, there is no reserve range where the curve (5) is perfectly flat, which would correspond to the region of abundant reserves. For inference on the transition between abundant and ample reserves, we rely on the IV estimates of the rate elasticity reported in Section 5, which suggest that this transition occurs when reserves are between 11% and 13% of banks' assets, depending on the time period.

θ_3 is a normalization factor that measures the vertical distance between the upper and lower asymptotes of the nonlinear time-invariant function in (5): $\lim_{x\to-\infty} f(x;\theta) - \lim_{x\to+\infty} f(x;\theta) = \pi\theta_3$. Reserves cannot go to minus infinity, but we can interpret this parameter in the limit of zero reserves. The theory predicts that, absent frictions, the federal funds rate should converge to the DW rate from below as reserves go to zero.[19] As a result, θ_3 should be of the same order of magnitude as the spread between the DW and IORB rates.

To ensure that minimizing (6) leads to economically meaningful results, we use the theoretical and institutional framework of Section 2 to initialize the parameters and set bounds on them. Appendix B 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 (6) 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 9 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 (3). 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 Panel (a) of 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 curve seems to move back toward its initial horizontal location by 1 pp, for a total shift to the right of 2 pp from the beginning to the end of the sample. The vertical location, in contrast, jumps further up in 2020-2021 by additional 10 bp, for a total increase of 12 bp relative to the first period. 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 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 raise in the federal funds rate cannot be interpreted as a signal of increased reserve scarcity. To identify the transition between abundant and scarce reserves, instead, one needs to

[19] Frictions such as stigma or borrowing caps may push the federal funds rate above the DW rate.

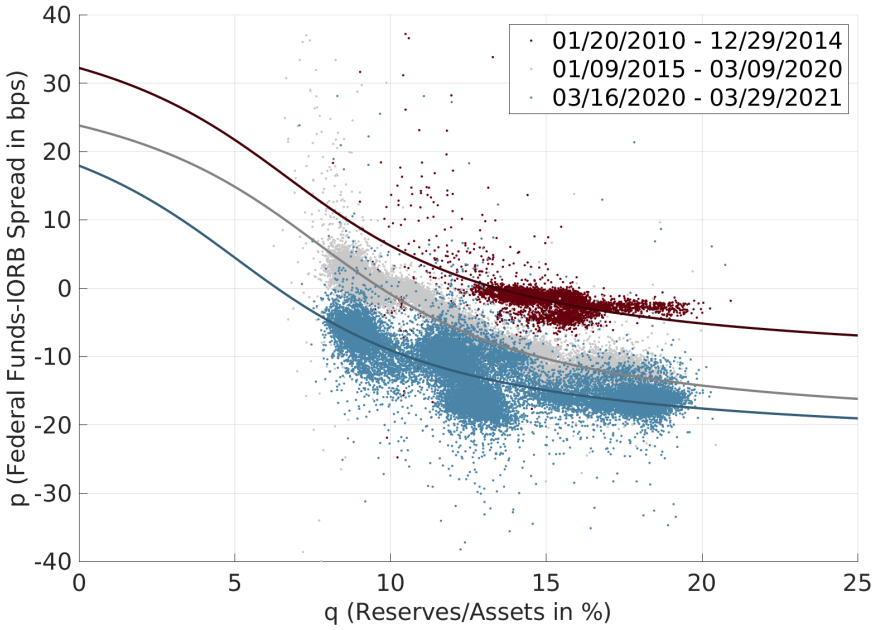


Figure 9: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data.** This figure shows the results of the nonlinear least-squares (NLLS) minimization in equation (6). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of the time-varying VAR model in equation (3). 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 (5). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B.

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 (θ_1) occurs when reserves are around 5% of banks' assets; the point of maximum slope growth ($\theta_1 + \theta_2/\sqrt{3}$), instead, is around 8%. This estimate suggests that, in 2010-2014, the transition between ample and scarce reserves occurred around 8% of banks' assets. In 2015-2020, 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 θ_3 is 18 bp, which is close to the average DW-IORB spread in our sample divided by π (15 bp), confirming that our results are reasonable.

Looking back at the path of realized reserves over 2010-2021, these results suggest that, in both 2010 and 2019, with reserves between 8% and 10% of banks' assets, the federal funds market may have been operating around the transition between ample and scarce reserves; in the second half of 2019, in particular, with reserves consistently below 9% of banks' assets and the threshold between

ample and scarce around 11%, the market may have been operating inside the scarcity region.

	1/2010-12/2014	01/2015-03/2020	03/2020-03/2021
	(a) Forecasts based on bivariate time-varying VAR		
Horizontal shifts q^*	0	2.62	1.69
Vertical shifts p^*	-23.67	-21.47	-11.94
	(b) Forecasts based on trivariate time-varying VAR with repo rates		
Horizontal shifts q^*	0	2.80	1.34
Vertical shifts p^*	-24.09	-22.03	-12.06
	(c) Forecasts based on trivariate time-varying VAR with treasury rates		
Horizontal shifts q^*	0	2.97	0.13
Vertical shifts p^*	-24.23	-22.48	-11.20

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 (5) from the nonlinear least-squares (NLLS) minimization in equation (6). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of federal funds-IORB spreads 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 (3) 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 (3). 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 (5). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B.

6.4 Robustness

For robustness, we repeat the nonlinear minimization (6) using the forecasts from the trivariate time-varying VAR with repo rates and with Treasury yields as pseudo-data. The coefficient estimates from these specifications are in Tables 2 and 3; Figure 11 and Figure 12 in Appendix B replicate Figure 9 for the trivariate models with repo rates and with Treasury yields, respectively. Results are consistent with those obtained using the bivariate forecasting model. The reserve demand curve shifts upward by 2 bp from 2010-2014 to 2015-2020 and by additional 10 to 11 bp in the last year of the sample, for a total vertical shift of 12 to 13 bp, which is substantial. The curve also shifts to the right by 3 pp from 2010-2014 to 2015-2020, but its horizontal position at the end of the sample is very close to the initial position: the total right shift is just 1 pp when using the model with repo rates and 0 pp when controlling for Treasury yields.

Forecasting model	θ_1 (%)	θ_2 (%)	θ_3 (bp)
Bivariate	5.02	5.31	17.87
Trivariate with repo rates	5.12	5.73	17.36
Trivariate with treasury rates	5.03	5.89	17.42

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 (5) from the nonlinear least-squares (NLLS) minimization in equation (6). 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 (3) 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 (3). 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 (5). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B.

For both trivariate models, the estimates of the parameters governing the nonlinear part of the demand curve are also similar to the baseline ones. In 2010-2014, the point of maximum slope is around 5% and the point of maximum slope growth around 8%. Combined with the evidence on the horizontal shifts, these results suggests that reserves would go from ample to scarce around 8% of banks' assets in 2010-2014 and around 11% in 2015-2020, confirming that in 2019, the federal funds market may have entered the scarcity region.

In Appendix B, we report additional robustness checks. For example, instead of fitting the curve to the right of the point of maximum slope, we fit it over a broader region that includes the point of maximum slope. We also repeat our post-processing exercise allowing the horizontal and vertical shifts to change at daily frequency; to do so, we include a loss function penalizing daily jumps in the path of p^* and q^* in the objective function (6). Results are similar and show that a large part of the time variation in the reserve demand curve is due to persistent upward shifts starting in 2013, whereas horizontal shifts seem to be less material.

7 Conclusion

Estimation of the reserve demand curve – the relationship between the prices at which banks are willing to trade their reserve balances and the aggregate reserves in the banking system – is important for the implementation of monetary policy. A key question is whether the price of reserves, the federal funds rate, responds to shocks to reserves even when reserves are in the trillions of dollars, as it has been since the 2008 financial crisis.

In this paper, we provide a structural time-varying estimate of the elasticity of federal funds rates to reserve shocks over 2010-2021, as reserves ranged from $1 to $4 trillion. We make several contributions. First, we make a methodological contribution in providing a methodology that can deal with the three main issues affecting the estimation of the reserve demand curve: nonlinearity, time variation due to slow-moving structural changes in the reserves market, and endogeneity. Our methodology uses of high-frequency (daily) data and an instrumental-variable approach combined with a time-varying vector autoregressive model of the joint dynamics of rates and reserves.

Second, we show that, as predicted by economic theory, the reserve demand curve is highly nonlinear: it is flat when reserves in the banking system are sufficiently large and negatively sloped as aggregate reserves decline. In the earlier part of the sample, we observe a significantly negative slope when reserves are below 12% of banks' assets (2010-2011); in the second half, a gentle slope re-emerges as normalized reserves drop below 13% (2018-2019).

Third, we show that the demand curve has shifted over time, both vertically and horizontally. 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 may not be the appropriate summary statistic for the rate sensitivity to reserve shocks.

Finally, while our focus is on estimating the curve's slope and documenting its shifts rather than explaining them, studying the causes of the shifts we document here is an important task for future research. Many factors may have moved banks' demand for reserves over the past decade. Lessons learnt from the 2008 crisis may have led to changes in banks' risk management and liquidity preferences. The introduction of liquidity regulation, supervisory stress tests, and resolution plans may have also played an important role. Structural changes in the reserves market, such as a drastic drop in inter-bank volumes, may have shifted the bargaining power across participants. Similarly, the introduction of the ONRRP facility may have impacted rates by providing an additional outside option to overnight non-bank lenders such as FHLBs and MMFs.

References

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. (2020a). The market events of mid-September 2019. *Federal Reserve Bank of New York Staff Reports No. 918*, 918.

Afonso, G., Kim, K., Martin, A., Nosal, E., Potter, S., and Schulhofer-Wohl, S. (2020b). Monetary policy implementation with an ample supply of reserves. *Federal Reserve Bank of New York Staff Reports No. 910*.

Anderson, A. G., Du, W., and Schlusche, B. (2020). Arbitrage capital of global banks. *NBER Working Paper 28658*.

Banegas, A. and Tase, M. (2020). Reserve balances, the federal funds market and arbitrage in the new regulatory framework. *Journal of Banking and Finance*.

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.

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.

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

Hamilton, J. D. (1997). Measuring the liquidity effect. *The American Economic Review*, 87(1):80–97.

Ihrig, J., Senyuz, Z., and Weinbach, G. (2020). The Fed's "ample-reserves" approach to implementing monetary policy. *Finance and Economic Discussion Series*.

Keister, T., Martin, A., and McAndrews, J. (2008). Divorcing money from monetary policy. *Federal Reserve Bank of New York Economic Policy Review*, 14:41–56.

Kim, K., Martin, A., and Nosal, E. (2020). Can the U.S. interbank market be revived? *Journal of Money, Credit and Banking*, 52:1645–1689.

Lopez-Salido, D. and Vissing-Jorgensen, A. (2022). Reserve demand and balance sheet run-off. *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. (2022). The financial market effects of unwinding the federal reserve's balance sheet. *Federal Reserve of Kansas City Working Paper*.

A Appendix: The Time-Varying VAR

A.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:

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

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

The vectorized form of model (A.1) is:

$$y_t = c_t + B_{1,t}y_{t-1} + ... + B_{m,t}y_{t-m} + u_t \qquad \text{with } t = 1, ..., T, \tag{A.2}$$

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

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

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

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \tag{A.3}$$

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

$$\begin{aligned} y_t &= c_t + B_{1,t}y_{t-1} + ... + B_{m,t}y_{t-m} + A_t^{-1}\Sigma_t \varepsilon_t, \\ Var(\varepsilon_t) &= I_n, \end{aligned} \tag{A.4}$$

where ε_t is a 2×1 vector of reserve and rate shocks that are uncorrelated with each other at each point in time by construction. The factorization of the covariance matrix in (A.3) is convenient because the first ε error is proportional to the forecast error in reserves ($\sigma_{1,t}\varepsilon_{1,t} = u_{q,t}$). As shown in the next section, this modeling strategy implies that the impulse response functions of q_t and p_t to $\varepsilon_{1,t-h}$ are proportional to the covariances of q_t and p_t with $u_{q,t-h}$.

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

$$y_t = X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \tag{A.5}$$
$$X_t' = I_n \otimes [1, y_{t-1}', \ldots, y_{t-m}'],$$

where $\otimes$ denotes the Kronecker product.

We model the parameters in the following way:

$$B_t = B_{t-1} + \nu_t, \tag{A.6}$$
$$\alpha_t = \alpha_{t-1} + \zeta_t, \tag{A.7}$$
$$\log \sigma_t = \log \sigma_{t-1} + \eta_t, \tag{A.8}$$

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

$$V = Var\left(\begin{bmatrix}\epsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t\end{bmatrix}\right) = \begin{bmatrix} I_2 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \tag{A.9}$$

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

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

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

so that α_t in (A.7) becomes a 3×1 vector of the stacked parameters of A_t. Σ_t maintains its diagonal structure and expands to include $\sigma_{3,t}$, so that σ_t in (A.8) becomes a 3×1 vector.

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

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

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

A.2 Covariance between errors and observables: an impulse-response view

In this section, we show how the covariances in equation (4) 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.3).

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

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

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

The Cholesky factorization (A.3) 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,

$$\operatorname{cov}\left(y_t, u_{1,t-h} | \Gamma_{1:T}\right) = \left(\mathbf{J}' \prod_{k=1}^{h} \mathbf{B}_{t-k} \mathbf{J} A_{t-h}^{-1} \iota_1\right) \sigma_{1t-h}^2. \tag{A.10}$$

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 A.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.10) are therefore equal to the h-day-ahead impulse responses of the system variables to the standardized reserve shocks calculated using factorization (A.3); in fact, in the traditional VAR literature, the i-th variable's impulse response to ε_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 (4), obtained as ratio of the covariances in (A.10), is exactly equal to the ratio of the h-day-ahead impulse responses of rates and reserves to reserve shocks.

A.3 Priors

We use Bayesian methods to estimate model (A.1). As outlined in Primiceri (2005), we use the following prior densities for the initial states of the time-varying parameters:

$$P(B_0) = N(\hat{B}, 4 \cdot \hat{\Psi}_B),$$
$$P(\alpha_0) = N(\hat{\alpha}, 4 \cdot \hat{\Psi}_\alpha),$$
$$P(\log \sigma_0) = N(\log \hat{\sigma}, I_n),$$

where $N(\mu, \sigma^2)$ denotes a normal density function with mean μ and variance σ^2, and $\hat{B}$, $\hat{\alpha}$, $\log \hat{\sigma}$, $\hat{\Psi}_B$, and $\hat{\Psi}_\alpha$ are set using a time-invariant VAR with the same ordering as in (A.3) estimated by OLS on the pre-sample from 01/05/2009 to 01/19/2010, covering $T_0 = 226$ daily observations. The prior means, $\hat{B}$ and $\hat{\alpha}$, are set to the OLS point estimates. The prior variances, $\hat{\Psi}_B$ and $\hat{\Psi}_\alpha$, are set equal to the sampling variances of the OLS point estimates. The prior means of the initial states of the log-volatilities are set to the logarithm of the standard errors of the OLS residuals.

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

$$P(Q) = IW(\lambda_1^2 \cdot T_0 \cdot \hat{\Psi}_B, T_0),$$
$$P(S) = IW(\lambda_2^2 \cdot 2 \cdot \hat{\Psi}_\alpha, 2),$$
$$P(W) = IW(\lambda_3^2 \cdot 3 \cdot I_3, 3),$$

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

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

A.4 Out-of-sample validation: forecasting

To evaluate the forecasting performance of the bi-variate TV-VAR model (A.1), we conduct a series of out-of-sample (OOS) forecasting exercises and compare the TV-VAR against traditional time-series models. 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 March 29, 2021, using an expanding window of observations.

For comparison, we also generate OOS forecasts using two time-invariant models: a standard bi-variate 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 five business days.

A.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 errors, 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 forecasts horizon. The only exception is the VAR's RMSE for the spread at $h = 10$, which is slightly smaller than the corresponding TV-VAR's RMSE: 2.08 bp vs 2.12 bp.

The TV-VAR displays a higher forecasting accuracy at all horizons also 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 large that their relationship with federal funds rates was completely flat, results are more mixed, and no model clearly dominates the others.

Sample	Model	Reserves/Assets (%)			FF-IORB Spread (bp)			Joint		
		h = 5	h = 10	h = 20	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
2010 - 2021	TVVAR	0.41	0.61	0.87	1.73	2.12	2.37	0.70	1.24	2.00
	VAR	0.44	0.65	0.99	2.04	2.08	2.69	0.88	1.30	2.43
	AR	0.43	0.63	1.01	2.74	3.99	6.37	-	-	-
2010 - 2011	TVVAR	0.35	0.46	0.68	1.47	1.76	2.23	0.45	0.74	1.43
	VAR	0.38	0.53	0.81	1.66	2.09	2.93	0.52	0.90	1.81
	AR	0.36	0.50	0.75	1.58	1.99	2.77	-	-	-
2012 - 2013	TVVAR	0.34	0.44	0.53	1.29	1.74	2.47	0.42	0.72	1.23
	VAR	0.35	0.47	0.65	1.38	1.87	2.70	0.46	0.80	1.49
	AR	0.35	0.46	0.61	1.37	1.90	2.69	-	-	-
2014 - 2015	TVVAR	0.58	0.84	1.04	0.78	1.04	1.26	0.46	0.87	1.32
	VAR	0.62	0.78	0.816	0.82	1.04	1.29	0.51	0.81	1.05
	AR	0.59	0.77	0.85	0.78	0.99	1.29	-	-	-
2016 - 2017	TVVAR	0.36	0.53	0.75	0.42	0.65	0.97	0.15	0.34	0.72
	VAR	0.38	0.60	0.83	0.47	0.72	0.99	0.17	0.41	0.78
	AR	0.37	0.56	0.77	0.46	0.71	1.04	-	-	-
2018 - 2019	TVVAR	0.26	0.37	0.47	2.45	2.79	2.98	0.55	0.82	1.13
	VAR	0.31	0.46	0.55	3.37	2.70	3.57	0.94	1.05	1.74
	AR	0.27	0.39	0.52	5.43	8.28	14.0	-	-	-
2020 - 2021	TVVAR	0.53	0.91	1.60	3.14	3.94	3.82	1.64	3.48	5.91
	VAR	0.55	1.03	2.10	3.19	3.49	3.96	1.75	3.57	7.19
	AR	0.57	1.07	2.23	3.32	4.02	4.04	-	-	-

Table 4: **OOS RMSE**. Out-of-sample (OOS) root-mean-square error (RMSE) for normalized reserves and the federal funds-IORB spread from the TV-VAR (A.1) and the VAR and AR models described in Section A.4.

A.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 the forecasting model. Lower scores correspond to lower likelihoods, suggesting lower model's 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, but the difference is practically immaterial at horizons of 5 and 10 business days, (e.g., for $h = 5$, the spread's log-score is -2.31 for the TV-VAR and -2.29 for the VAR). 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 the relationship between rates and reserves was practically zero, that the TV-VAR's performance tends to be slightly worse.

Sample	Model	Reserves/Assets (%)			FF-IORB Spread (bp)			Joint		
		h = 5	h = 10	h = 20	h = 5	h = 10	h = 20	h = 5	h = 10	h = 20
2010 - 2021	TVVAR	-0.51	-0.87	-1.38	-2.31	-2.47	-3.38	-2.92	-3.34	-4.72
	VAR	-0.69	-1.17	-1.97	-2.29	-2.34	-2.71	-3.21	-3.79	-5.04
	AR	-0.60	-1.08	-1.83	-2.28	-2.39	-2.62	-	-	-
2010 - 2011	TVVAR	-0.39	-0.67	-1.13	-1.82	-2.02	-2.29	-2.09	-2.64	-3.35
	VAR	-0.58	-0.96	-1.77	-2.02	-2.29	-2.76	-2.40	-3.09	-4.44
	AR	-0.50	-0.86	-1.47	-1.94	-2.19	-2.60	-	-	-
2012 - 2013	TVVAR	-0.37	-0.66	-0.97	-1.87	-2.18	-2.69	-2.24	-2.83	-3.61
	VAR	-0.43	-0.82	-1.31	-1.75	-2.13	-2.70	-2.13	-2.88	-3.83
	AR	-0.41	-0.74	-1.09	-1.74	-2.14	-2.67	-	-	-
2014 - 2015	TVVAR	-1.11	-1.45	-1.61	-1.62	-2.45	-4.25	-2.75	-3.93	-5.80
	VAR	-1.25	-1.41	-1.43	-1.30	-1.59	-1.91	-2.64	-3.17	-3.43
	AR	-1.11	-1.39	-1.49	-1.21	-1.47	-1.86	-	-	-
2016 - 2017	TVVAR	-0.43	-0.83	-1.54	-1.08	-2.22	-4.23	-1.51	-3.05	-5.72
	VAR	-0.46	-0.93	-1.30	-0.55	-0.90	-1.32	-1.03	-1.92	-2.80
	AR	-0.42	-0.82	-1.14	-0.55	-0.89	-1.31	-	-	-
2018 - 2019	TVVAR	-0.16	-0.53	-0.94	-5.16	-3.50	-3.31	-5.88	-4.03	-4.22
	VAR	-0.35	-0.78	-0.96	-5.56	-4.55	-4.75	-6.96	-6.18	-6.67
	AR	-0.13	-0.51	-0.85	-5.65	-4.90	-4.46	-	-	-
2020 - 2021	TVVAR	-0.62	-1.14	-2.45	-2.19	-2.38	-3.37	-2.96	-3.56	-5.83
	VAR	-1.24	-2.67	-6.86	-2.68	-2.74	-2.87	-4.44	-6.38	-11.3
	AR	-1.26	-2.74	-6.67	-2.73	-2.94	-2.95	-	-	-

Table 5: **Mean Log Scores**: Average of the marginal log scores for Reserves and Spread (VWA FFR - IOR), and joint log scores for from the TV-VAR (A.1) and the VAR and AR models described in Section A.4.

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 $[0, 1]$. Figure10 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.

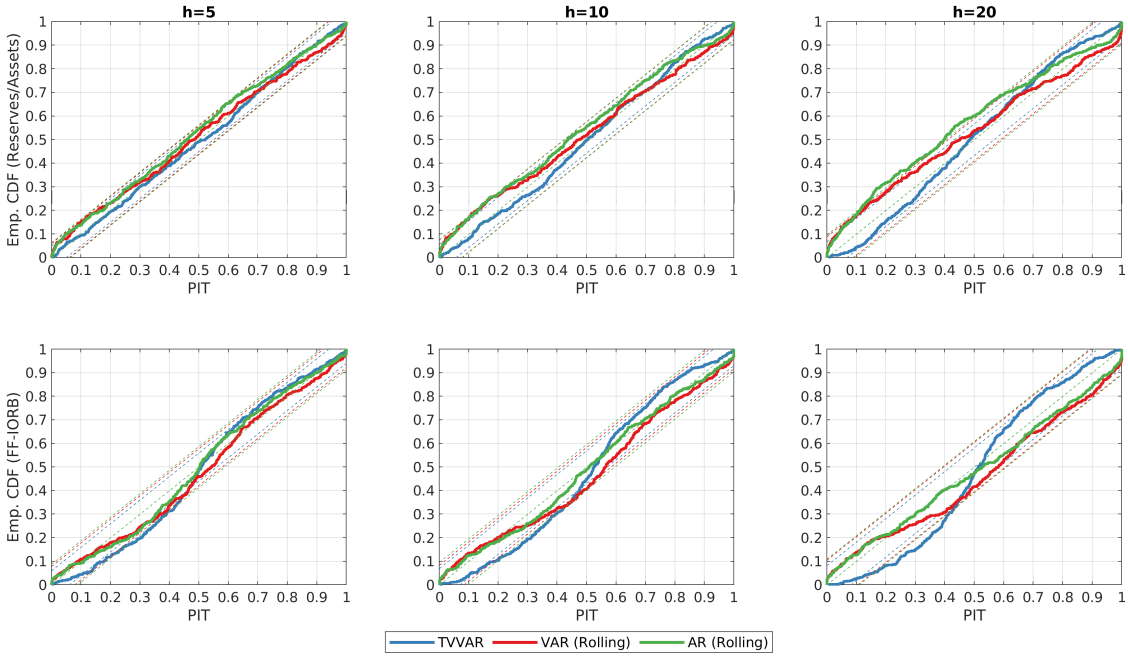


Figure 10: **Empirical CDFs of PITs**. Empirical cumulative distribution function (CDF) of the probability integral transforms (PITs) of the forecasts of normalized reserves and of the federal-funds-IORB spread from the TV-VAR (A.1) and the VAR and AR models described in Section A.4. The colored dashed lines represent the 90% confidence bands for the different models, constructed using the bootstrap method outlined in Rossi and Sekhposyan (2019)

B Appendix: The Post-processing Nonlinear Fit

B.1 Algorithm details

To solve the minimization problem in (6), 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 (6) and include it in the minimization program.

To ensure that our results are reliable and economically meaningful, we set bounds on the variables of the minimization problem. We are agnostic about the origins and possible magnitude of the horizontal shift, q^*, in the reserve demand curve; for this reason, in all periods, we set its upper and lower bounds equal to two and minus two times the maximum level of normalized reserves in our sample, respectively. For the vertical shift p^*, we choose bounds based on the discussion of its economic interpretation in 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 (5) is strictly positive everywhere, which implies that $p_t^* < p_t$ for all t by construction.

For the θ parameters governing the nonlinearity of the demand curve in (5), we impose the following bounds. θ_1 represents the point of maximum absolute slope of the demand curve, where reserves are highly scarce. To bound θ_1 from below, we calculate the ratio between the aggregate reserve requirement and banks' total assets on each day, compute the minimum value in our sample, and then take 10% of that value. To bound θ_1 from above, we multiply the maximum of normalized reserves in our sample by two. We use the same bounds for θ_2, which measures the distance between the point of maximum absolute slope and the point of maximum slope growth (i.e., the transition point between scarce and ample reserves) in the nonlinear function in (5).

The upper asymptote of the nonlinear function (5) is equal to $p_t^* + \pi\theta_3$. In the federal funds market, absent stigma, frictions, and caps on discount-window (DW) borrowing, the federal funds rate should be bounded from above by the DW rate. Therefore, given our bound on p^*, we set the upper bound on θ_3 to be equal to the maximum of the spread between the DW and ONRRP rates in our sample divided by π. The lower bound on θ_3 is simply set to zero.

To initialize the variables, we also build on our discussion of their economic interpretation in Section 6. We initialize θ_1 with the average level of reserves in 2009, which is the period of lowest reserve balances since the 2008 crisis; reserves in 2009 were most likely scarcer than in the rest of our sample. To initialize θ_2, we exploit the experience of September 2019, which suggests that reserves may have transitioned from ample to scarce around that time (Afonso *et al.*, 2020a). Since the point of maximum slope growth in (5) is $\theta_1 + \theta_2/\sqrt{3}$, we initialize θ_2 with $\sqrt{3}(q_{2019} - \theta_1^{(0)})$, where q_{2019} is the average value of reserves in September 2019, and $\theta_1^{(0)}$ is the initialization of θ_1.

The initial value of θ_3 is set equal to the average spread between the DW and IORB rates in our sample divided by π.

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
$\theta_1 \in [0.1\underline{q}^{req}, 2\bar{q}]$	0.18	37.17	$\theta_1^{(0)} = q_{2009} = 7.3$
$\theta_2 \in [0.1\underline{q}^{req}, 2\bar{q}]$	0.18	37.17	$\theta_2^{(0)} = \sqrt{3}(q_{2019} - \theta_1^{(0)}) = 2.89$
$\theta_3 \in [0, \pi^{-1} \max(\text{DW}_t - \text{ONRRP}_t)]$	0	25.5	$\theta_3^{(0)} = \pi^{-1}\text{avg}(\text{DW}_t - \text{IORB}_t) = 15.3$
$q_t^* \in [-2\bar{q}, 2\bar{q}]$	-37.17	37.17	$q_t^{*(0)} = 0$ for all t
$p_t^* \in [\min(\text{ONRRP}_t - \text{IORB}_t), \text{avg}(p_t)]$			$p_t^{*(0)} = \min(p_t) - 1$
01/20/2010-12/29/2014	-25	-12.52	-20.00
01/09/2015-03/09/2020	-25	-6.33	-15.00
03/16/2020-03/29/2021	-25	-1.73	-5.88

Table 6: **Bounds and initializations of the variables in the minimization of the objective function (6)**. θ_1, θ_2, and θ_3 are the parameters defining the nonlinear time-invariant functional form of the reserve demand function in (5); q^* and p^* are the horizontal and vertical shifts. $\underline{q}^{req}$ is the in-sample minimum of the aggregate required reserves normalized by banks' total assets; $\bar{q}$ is the in-sample maximum of normalized reserves; q_{2009} and q_{2019} are the average levels of reserves normalized by banks' assets in 2009 and in September 2019; DW is the discount-window rate; IORB is the interest rate on reserve balances; and ONRRP is the overnight reverse repurchase agreement rate.

B.2 Robustness

B.2.1 Controlling for repo rates and Treasury yields

Figure 11 shows the results of the nonlinear fit (6) on the joint forecasts of the federal funds-IORB spread and aggregate reserves from the trivariate time-varying VAR model that includes repo rates as additional variable.[20] Our findings are similar to those obtained using forecasts from the baseline bivariate model (see Figure 9). There seems to be a modest shift to the right from 2010-2014 to 2015-2020, but the most significant structural movements in the curve are upward shifts, especially in the last part of the sample (see Tables 2 and 3 in Section 6).

[20]Details on the specification of these trivariate models and their priors are in Appendix A.1.

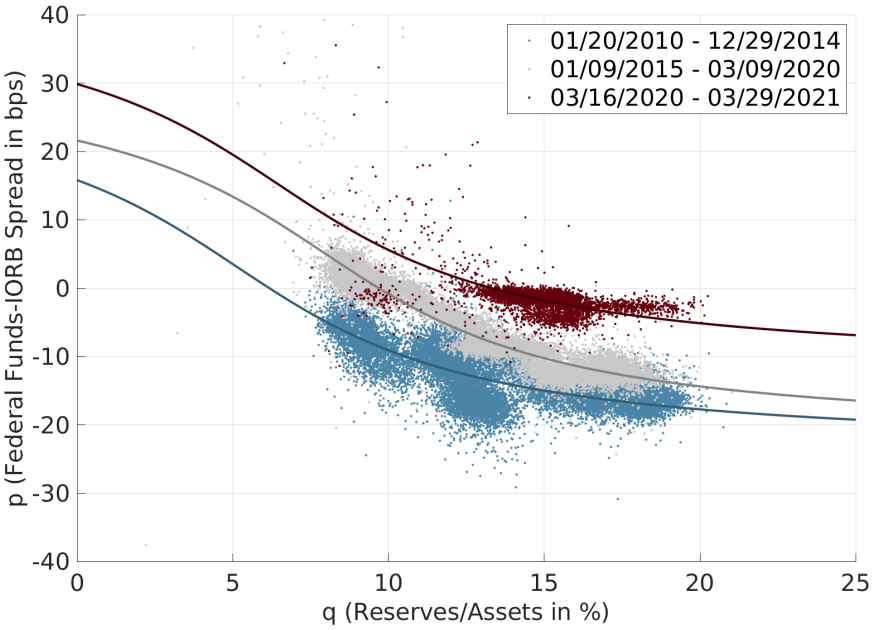


Figure 11: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data.** This figure shows the results of the nonlinear least-squares (NLLS) minimization in equation (6). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of the time-varying VAR model in equation (3). 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 (5). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B.

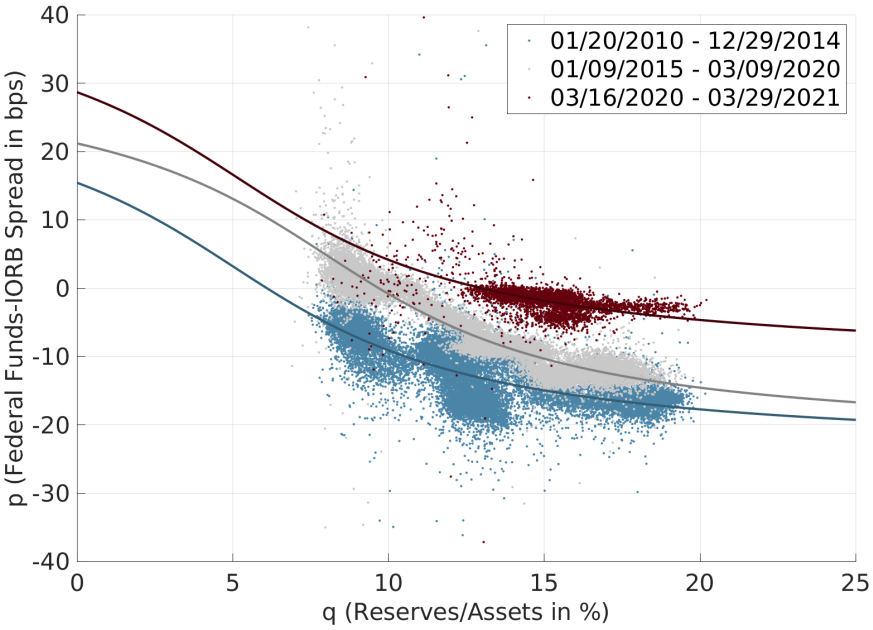


Figure 12: **Post-processing nonlinear fit of the reserve demand curve with horizontal and vertical shifts using model forecasts as data.** This figure shows the result of the nonlinear least-squares (NLLS) minimization in equation (6). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of the federal funds-IORB spread and normalized reserves from the in-sample estimation of the time-varying VAR model in equation (3). 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 (5). A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B.

B.2.2 Fitting the curve on a larger range

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 the following constraint: $q_{it} - q_t^* > \theta_1 - \theta_2/\sqrt{3}$; it is easy to show that, in fact, $\theta_1 - \theta_2/\sqrt{3}$ is the point at which the absolute slope of the curve decreases at the fastest rate as reserves decrease. This is a milder constraint as this region includes the point of maximum slope.

Results 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 12-13 bp and a more modest horizontal shift to the right of 1 to 2 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	03/2020-03/2021
	(a) Forecasts based on bivariate time-varying VAR		
Horizontal shifts q^*	0	2.75	1.69
Vertical shifts p^*	-21.10	-19.00	-9.23
	(b) Forecasts based on trivariate time-varying VAR with repo rates		
Horizontal shifts q^*	0	2.86	1.12
Vertical shifts p^*	-21.16	-19.13	-8.77
	(c) Forecasts based on trivariate time-varying VAR with treasury rates		
Horizontal shifts q^*	0	2.93	0.82
Vertical shifts p^*	-21.07	-19.19	-8.48

Table 7: **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 (5) from the nonlinear least-squares (NLLS) minimization in equation (6). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of federal funds-IORB spreads 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 (3) to generate the forecasts; results in panel (b) are obtained using the trivariate model that adds repo rates to model (3). Forecasts are generated every five days; for each day, we draw $N = 100$ forecasts from the model-implied posterior joint distribution. A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B

Forecasting model	θ_1 (%)	θ_2 (%)	θ_3 (bp)
Bivariate	7.45	4.42	11.51
Trivaraite with repo rates	7.63	4.47	11.17
Trivaraite with treasury rates	7.64	4.46	11.07

Table 8: **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 (5) from the nonlinear least-squares (NLLS) minimization in equation (6). 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 (3) to generate the forecasts; results in panel (b) are obtained using the trivariate model that adds repo rates to model (3). Forecasts are generated every five days; for each day, we draw $N = 100$ forecasts from the model-implied posterior joint distribution. A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B

B.3 Post-processing NLLS with daily shifts

In this section, we present a modification of our post-processing exercise in which we let the horizontal and vertical shifts change at daily frequency, instead of fixing them within three periods spanning several years (i.e., 2010-2014, 2015-3/2020, and 3/2020-3/2021 periods). To do so, we re-write the objective function (6) as follows

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\left[(p_{it}-p_t^*)-f(q_{it}-q_t^*;\theta)\right]^2+N\lambda_p\sum_{t=2}^{T-1}\left[\Delta^2 p_{t+1}^*\right]^2+N\lambda_q\sum_{t=2}^{T-1}\left[\Delta^2 q_{t+1}^*\right]^2 \tag{A.11}$$

where N is the number of draws from our joint forecast distribution ($N = 100$), λ_p and λ_q are positive constants, $\Delta^2 p_{t+1}^* = \left(p_{t+1}^* - p_t^*\right) - \left(p_t^* - p_{t-1}^*\right)$, $\Delta^2 q_{t+1}^* = \left(q_{t+1}^* - q_t^*\right) - \left(q_t^* - q_{t-1}^*\right)$, and all other variables are defined as in Section 6. The goal is to find $\left\{\hat{\theta}, \hat{p}_t^*, \hat{q}_t^*; t = 1, ..., T\right\}$ that minimize (A.11).

The idea of this post-processing exercise is to perform a nonlinear least-squares estimation with horizontal and vertical shifts at daily frequency that imposes that these shifts are slow-moving. By including terms that are proportional to the sum of the squares of the double-differences in the shifts, the objective function (A.11) penalizes sudden jumps and fast growth in p^* and q^*. The larger are the constants λ_p and λ_q, the higher is the penalty.

Since the minimization problem in (A.11) is somewhat similar to the Hodrick-Prescott filter commonly used in macroeconomics, we follow the same procedure to set the lambdas; namely, since

our forecast are at the daily frequency (i.e., we use forecasts every five days), we set $\lambda_p = \lambda_1 = (52 * 10)^2$. For robustness, we also considered alternative specifications, ranging from $(52 * 1)^2$ to $(52 * 20)^2$, and obtain qualitatively similar results.

The results of this nonlinear fit with daily-varying shifts are in Figures 13, 14, 15, and in Table 9. Consistent with the results in Section 6, these results suggest that vertical shifts seem to be more important than horizontal ones in explaining the movements of the reserve demand curve over time. In particular, starting from late 2013, the lower bound of the curve, p^*, seems to steadily move upward, implying a vertical shift of around 15 bp from the beginning to the end of the sample.

In terms of the time-invariant nonlinear part of the curve, the point of maximum absolute slope is attained when aggregate reserves are around 8% of banks' assets, and the point of maximum slope growth, which we interpret as the transition between ample and scarce reserves, is around 10%. These numbers are consistent with our results in Section 6 and suggest that, in 2019 (as in 2010), as reserves dropped below 10% of banks' assets, reserves may have transitioned from being ample to being scarce.

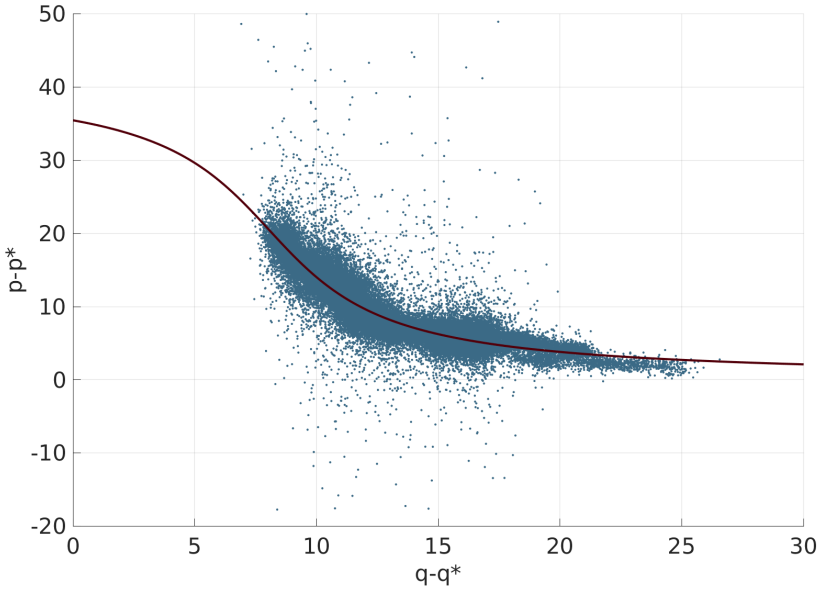


Figure 13: **Post-processing fit of nonlinear reserve demand function with horizontal and vertical shifts on bivariate model forecasts.** This figure shows the results of the nonlinear least-squares (NLLS) minimization in equation (A.11). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of federal funds-IORB spreads and normalized reserves from the in-sample estimation of the TV-VAR model in equation (3). 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 (sigmoid-shaped) demand function. A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in appendix.

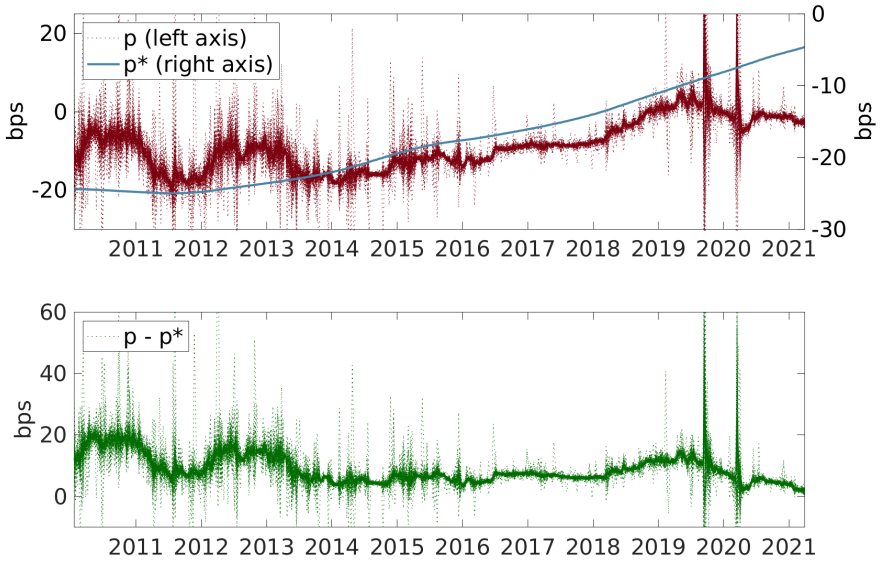


Figure 14: **Post-processing fit of nonlinear reserve demand function with horizontal and vertical shifts on bivariate model forecasts: time path of vertical shifts p^*.**

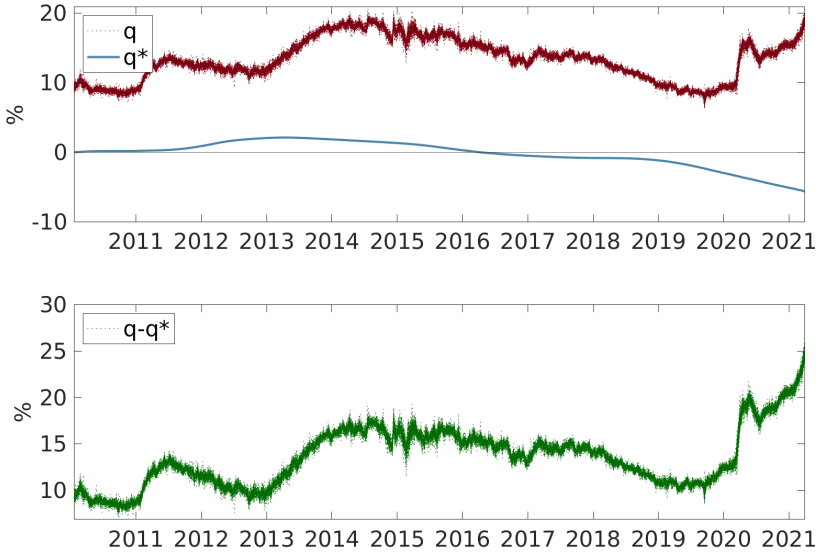


Figure 15: **Post-processing fit of nonlinear reserve demand function with horizontal and vertical shifts on bivariate model forecasts: time path of horizontal shifts q^*.**

Forecasting model	θ_1 (%)	θ_2 (%)	θ_3 (bp)
Bivariate	8.05	3.61	13.04
Trivariate with repo rates	7.93	3.84	12.35
Trivariate with treasury rates	8.04	3.69	12.82

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.** This table shows the estimates of the horizontal (q^*) and vertical (p^*) shifts in equation (5) from the nonlinear least-squares (NLLS) minimization in equation (6). The NLLS fit is estimated on a sample of five-day-ahead joint forecasts of federal funds-IORB spreads 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 (3) to generate the forecasts; results in panel (b) are obtained using the trivariate model that adds repo rates to model (3). Forecasts are generated every five days; for each day, we draw $N = 100$ forecasts from the model-implied posterior joint distribution. A detailed description of the minimization algorithm, parameter bounds, and initialization values can be found in Appendix B.