Brookings Papers
ON ECONOMIC ACTIVITY

BPEA Conference Draft, March 26-27, 2026

The Payment System Puts a Floor on the Fed's Balance Sheet

Darrell Duffie (Stanford Graduate School of Business)

The Payment System Puts a Floor on the Fed's Balance Sheet

Conference Draft

Brookings Papers on Economic Activity

Darrell Duffie
GSB Stanford

March 18, 2026*

Abstract

The quantity of deposits of banks at the Fed, known as reserve balances, is occasionally insufficient for the largest banks to make timely payments to each other. When that happens, money-market rates can become volatile or spike significantly. To prevent these outcomes, the Fed is now growing its balance sheet roughly in step with the quantity of reserve balances needed by the payment system. If the Fed were to aim for a smaller balance sheet, subject to effective monetary policy implementation, policies to consider include counteracting unintended shocks to the supply of reserve balances with temporary open market operations, adding a liquidity savings mechanism to the Fed's largest payment system, changing liquidity regulations, and tiering the remuneration of reserves.

Keywords: Payments, reserve balances, monetary policy implementation

JEL codes: E42

*I am grateful for related research collaborations with Gara Afonso, Adam Copeland, Peter Glynn, Reina Li, Lorenzo Rigon, Hyun Shin, Srisht Fateh Singh, Thanawat Sornwanee, Francesco Spizzuoco, Andreas Veneris, Chaojun Wang, Sarah Yu Wang, and Yilin Yang. I am also grateful for data from Lou Crandall, Jay Kahn, and Farooq Akram; research assistance from Rafael Berriel, Andrew Caosun, and Cynthia Wang; helpful conversations with Garth Baughman, Morten Bech, Markus Brunnermeier, John Cochrane, Peter Fisher, Rodney Garratt, Ben Hébert, Jay Kahn, Lorie Logan, Antoine Martin, Matthew McCormick, Seth Searls, Matt Zames, and Haoxiang Zhu; and comments on draft versions by Wenxin Du, Janice Eberly, Toni Gravelle, Bob Hodrick, Frank Keane, Arvind Krishnamurthy, Antoine Lallour, Jamie McAndrews, Brian Madigan, Dina Marchioni, Bill Nelson, Randall Quarles, Ricardo Reis, Brian Sack, Sam Schulhofer-Wohl, Matt Roberts-Sklar, Jón Steinsson, Anuj Tandon, Dan Tarullo, Colin Teichholtz, Annette Vissing-Jorgensen, Chris Waller, and several anonymous commenters.

Contents

1 Introduction

The quantity of deposits of banks at the Fed — known as reserve balances, or just "reserves" — is about $3 trillion, roughly 200 times pre-2008 levels. Even so, reducing the supply of reserve balances could lead some large banks to delay their payments to each other. When that happened in mid-September 2019, key money-market rates suddenly jumped far above the Fed's target. To prevent these outcomes, the Fed is now growing its balance sheet roughly in step with the quantity of reserve balances needed by the payment system. If the Fed were to target a smaller balance sheet, subject to effective monetary policy implementation, this paper explains a range of policies that the Fed could consider, including counteracting unintended shocks to the supply of reserve balances with temporary open market operations, adding a liquidity savings mechanism to the Fed's largest payment system, changing liquidity regulations, and tiering the remuneration of reserves.

Under liquidity regulations introduced after the Global Financial Crisis (GFC), systemically important bank holding companies (GSIBs) must demonstrate that their liquidity needs can be met from their own resources. This reduces their incentives to obtain liquidity from the Fed, whether through daylight overdrafts, the Discount Window, or Standing Repo Operations (SRPs). These sources of liquidity are stigmatized (Nelson, 2026a). At some points in time since 2018, the Fed has reduced system-wide reserve balances to the point that some large banks conserve on their reserve balances by throttling their outgoing payments, with adverse self-fulfilling equilibrium expectations of payment delays. In these situations, the constraint of having sufficient reserve balances to make timely outgoing payments throughout the day without relying on funding from the Fed can suddenly have a high and volatile shadow price, disrupting the Fed's targeting of market interest rates. It has been difficult for the Fed to accurately predict the supply of reserves necessary to avoid this outcome.

As a sign of this mechanism at work, on October 29, 2025, the Federal Open Market Committee announced that it would stop reducing the size of the Fed's balance sheet. This decision was triggered by the increasing level of wholesale market interest rates relative to the interest rate paid to banks by the Fed on their reserve balances (IORB), as shown in Figure 1. For example, on October 31, 2025, the Secured Overnight Financing Rate (SOFR) was 32 basis points above IORB. (SOFR is the volume-weighted median rate of approximately $3 trillion of overnight Treasury repo transactions.[1]) As a natural next step, on December 10, 2025, the Fed relaxed some constraints on its Standing Repo Operations and began a program of "reserve management purchases" (RMPs), buying Treasury bills to create about $40 billion of additional reserves each month until May 2026.[2]

[1]Data are sourced from FRED for IORB, Federal Reserve Bank of New York for SOFR volume and rate data.

[2]See Federal Reserve Bank of New York (2025). On December 10, 2025, at his Press Conference, Fed

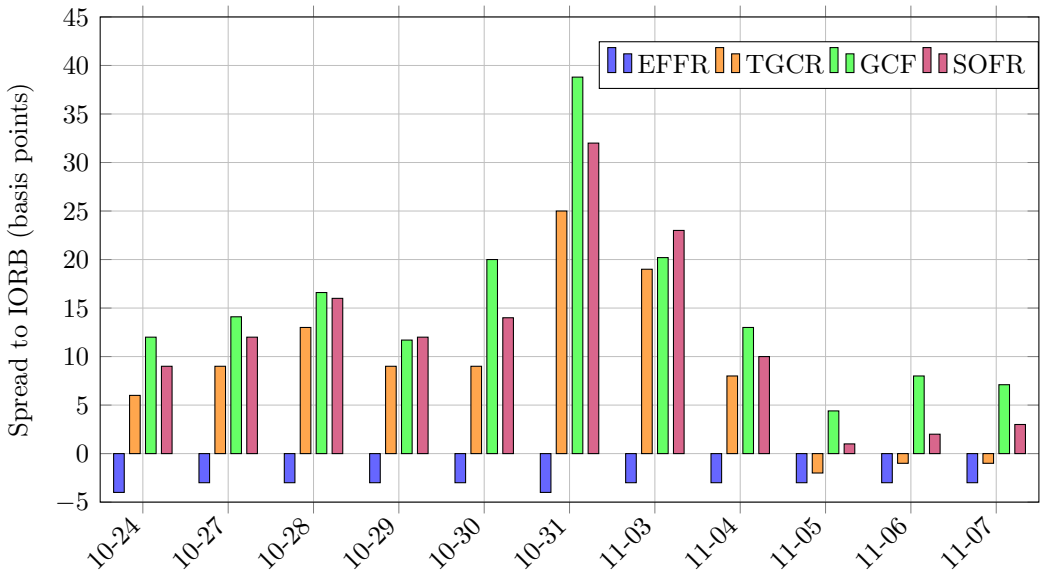


Figure 1: Select reference rates, spread to IORB, in the days surrounding October 31, 2025. Effective federal funds rate (EFFR), Tri-party General Collateral Rate (TGCR), GCF Repo® (GCF), Secured Overnight Financing Rate (SOFR). The SRF rate during this period was 10 basis points above IORB. Data: FRED, Federal Reserve Bank of New York, DTCC.

Section 2 offers an overview of the Fed's policy choices, including a discussion of the tradeoffs involving the size of its balance sheet without arriving at any cost-benefit conclusion. If the Fed does decide to go for a smaller balance, some enabling policy options are summarized here. Section 3 describes the workings of the Fed's main payment systems and their role in determining the demand for reserve balances. Section 4 outlines how the Fed's framework for monetary policy implementation has evolved since the GFC, depending on whether reserves are scarce, abundant, or "ample," and challenges associated with attaining the minimum ample level of reserves.

The remainder of the paper focuses on four policies that would reduce the required supply of reserve balances without disrupting monetary policy implementation. The first of these, described in Section 5 would use temporary open market operations to offset unintended shocks to the supply of reserve balances, allowing a smoother and lower average path for the supply of reserves. Section 6 considers the role of liquidity regulation and supervision. Section 7 explores how the Fed could generate some of the strong pre-GFC incentives for banks to lend their excess reserves by "tiering down" the interest rate that it pays on ex-

Chair Jay Powell said: "we also decided to initiate purchases of shorter-term Treasury securities solely for the purpose of maintaining an ample supply of reserves over time, thus supporting effective control of our policy rate." And, in answer to a question, Powell said "So, you know, we knew this was going to come. When it finally did come, it came a little quicker than expected, but we were, yeah, absolutely there to take the actions that we said we would take. And those actions are today. So, you know, we announced that we're resuming reserve management purchases. That is completely separate from monetary policy. It'a just we need to keep an ample supply of reserves out there."

cess reserves (Armenter, 2016; Baughman and Carapella, 2019, 2020; Duffie, Sornwanee, and Spizzuocco, 2026). Section 8 discusses how the demand for reserve balances could be reduced by augmenting the Fed's largest payment system, Fedwire, with a liquidity savings mechanism that allows banks to make large outgoing payments with incoming payments, rather than by making payments out of "pre-loaded" reserve balances.

This paper does not focus on the fiscal implications of a reduction in the size of the Fed's balance sheet. If the Fed decides to back its reserve balances with Treasury bills, the size of the Fed's balance sheet will have relatively little fiscal impact. However, to the extent that the Fed backs its reserve balances with longer-term Treasuries and eventually obtains a substantial reduction in the supply of reserve balances, private investors would need to absorb more long-term Treasuries. The results of Krishnamurthy and Vissing-Jorgensen (2011) imply that the federal government would in that case pay a correspondingly higher average interest rate on its Treasury debt.

2 Policy Overview

This section provides some background on the liability side of the Fed's balance sheet, describes the importance of monetary policy transmission into the Treasury repo market, and offers an overview of the Fed's policy options for reducing the size of its balance sheet.

2.1 Reserve balances: supply driven or demand driven?

As shown in Figure 2, the second largest liability of the Fed is paper currency, whose quantity usually adjusts slowly and predictably over time and because of that does not disrupt the Fed's monetary policy implementation.[3] The next largest liability of the Fed is the U.S. Treasury Department's General Account (TGA) at the Fed, whose balance is volatile and does play a role in determining the size of the Fed's balance sheet (Vissing-Jorgensen, 2025; Duygan-Bump and Kahn, 2026). However, the role of the TGA is through the same reserve-balances channel that I described. That is, when TGA balances go up, the quantity of reserve balances goes down one for one, mechanically. (Virtually all payments to the U.S. Treasury are made by crediting its TGA account and debiting the Federal Reserve account of the payee's bank. Payments by the U.S. Treasury are done by the reverse credit and debit.) If, as a result of increases in the TGA, there are not enough reserve balances left to run the payment system effectively, money market rates are disrupted. A high or rising TGA, on its own, plays no direct role in money markets if the Fed ensures that reserve balances remain

[3]Exceptionally, in March and April of 2020 the COVID pandemic caused an unusually large expansion of the supply of paper currency, by $94 billion (FRED).

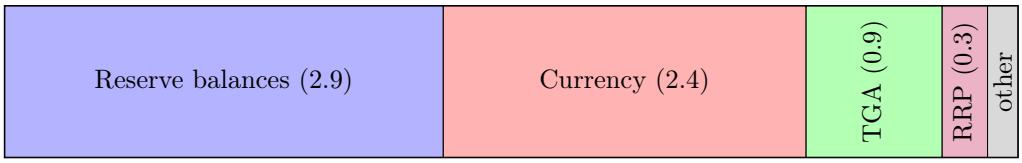


Figure 2: The liabilities of the Federal Reserve System in trillions of dollars, to scale, as of February 5, 2026. TGA refers to the Treasury General Account. RRP refers to the sum of the Foreign Official Repo Facility ($320 billion) and the Overnight Reverse Repurchase Facility ($6 billion). Data: Federal Reserve Board, H.4.1.

sufficient, if necessary by offsetting TGA increases with temporary open market operations that create additional reserves (Vissing-Jorgensen, 2025; Nelson, 2024).

Whenever reserve balances are more than ample, the Fed has purposefully allowed its asset portfolio to decline, extinguishing reserves in tandem, unless new asset purchases are needed for a specific objective such as quantitative easing or mitigating dysfunction in Treasury markets. This process is called "balance sheet normalization" or "quantitative tightening." Other things equal, that is, the Fed prefers a smaller balance sheet. The Fed has been using elements of two basic approaches to achieving the smallest balance sheet consistent with "ample" reserves:

- **Supply driven.** Maintain a supply of reserves that is conservatively sufficient to meet the needs of the payment system, even assuming that large financial institutions are reluctant to obtain additional reserves from the Fed's liquidity facilities.

- **Demand driven.** Provide a somewhat smaller standing supply of reserves, small enough that there would be episodes of significantly elevated repo rate spreads unless financial institutions obtain temporary funding from the Fed's liquidity facilities.

As I will explain, the Bank of England has had some success with the demand-driven approach (Saporta, 2025; Dolan and Roberts-Sklar, 2025), leading to a substantial reduction in the total supply of sterling reserves, even when significant amounts of funding are obtained from the Bank of England's Short Term Repo operations. Although the European Central Bank still has abundant reserve balances, it plans for its ongoing reduction of its balance sheet to be managed with a demand-driven approach to providing reserves. In November 2025, ECB Executive Board member Isabel Schnabel stated:[4]

In a demand-driven framework, the central bank's operations are not a liquidity backstop. They are there to be used by banks as part of their day-to-day liquidity management, which is also acknowledged by banking supervision. Our framework also relies on large dealer banks

[4]See Schnabel (2025).

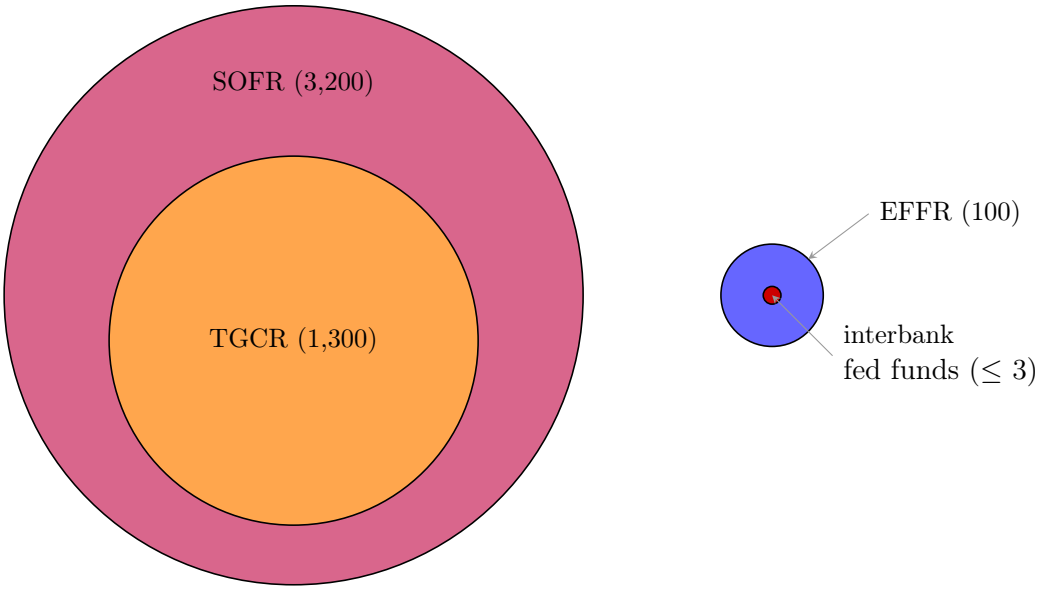


Figure 3: A Venn diagram showing, to scale, the approximate recent sizes of selected wholesale money-market segments. Daily volumes of transactions are shown in billions of dollars. The associated rate benchmarks are Secured Overnight Financing Rate (SOFR), the Tri-Party General Collateral Repo Rate (TGCR), the effective federal funds rate (EFFR). The transactions underlying the TGCR are a subset of those underlying SOFR. Data: SOFR and EFFR: Federal Reserve Bank of New York; EFFR: Office of Financial Research. The total size of the Treasury repo market is approximately $8.8 trillion (Cenicola, 2025). Anbil, Anderson, and Eyal (2026) show that recent daily interbank federal funds loan volumes have ranged from 2 to 3 billion dollars.

supporting financial intermediation. The more connected these banks are to our operations, and the more frequently they use them, the less likely it is that repo rates will spike, as dealer banks often act as liquidity hubs within countries and banking networks.

2.2 Focus on transmission of policy into repo rates

When characterizing the implementation of monetary policy in market interest rates, I will focus on rates set in the Treasury repo market, rather than adopt the Fed's traditional focus on the effective federal funds rate (EFFR). The Treasury repo market is a far more significant money market, with approximately $8.8 trillion in outstanding transactions (Cenicola, 2025), most of which are daily.[5] By contrast, the total daily volume in the federal funds market was only about $100 billion in February 2026 (FRED), with less than $3 billion of this total in the form of lending between banks (Anbil, Anderson, and Eyal, 2026). Figure 3 compares interbank loan volumes in the federal funds market with the much larger volumes underlying two Treasury repo rate benchmarks, the Secured Overnight Financing Rate (SOFR) and the Tri-Party General Collateral Rate (TGCR). Most lending in

[5]This estimate is upward biased because transactions that are between two dealers are reported by both dealers, and thus doublecounted.

the federal funds market is between Federal Home Loan Banks, which receive no interest on their reserve balances, and foreign banks, which do not pay FDIC insurance fees.[6] The rates negotiated between these lenders and borrowers have limited relevance for wholesale U.S. money markets, particularly the enormous repo markets, which are critically important for the Fed's monetary policy transmission. The EFFR often fails to reflect stresses in the far more important repo market, as shown for example in Figure 1. The estimated demand elasticity of EFFR with respect to the supply of reserve balances, reported on an ongoing basis by the Federal Reserve Bank of New York,[7] has remained very low through periods of somewhat elevated and volatile Treasury repo rates associated with reduced supplies of reserve balances. Logan and Schulhofer-Wohl (2025) suggests that TGCR may be more appropriate than EEFR as the FOMC's monetary policy target.

Market repo rates are especially likely to be elevated on quarter ends and certain other "period-end" days on which a significant subset of foreign banks are monitored for capital adequacy and therefore slash their holdings of reserves and their repo funding to other market participants (Kloks, Mattille, and Ranaldo, 2025). On quarter ends, as shown in Figure 4, there have been significant corresponding drops in the supply of reserves, with large cash investors resorting to temporary investments in the Fed's Overnight Reverse Repurchase Facility. Analysis by Lou Crandall of Wrightson Associates on February 16, 2026 shows that since the second quarter of 2021 the total quarter-end "compression" of reserve balances by foreign branch offices (FBOs) has been at least $200 billion on 15 quarter ends, at least $300 billion on seven quarter ends, and at least $400 billion on two quarter ends. (A given reduction in FBO reserve balances does not necessarily imply the same reduction in system-wide balances.) At the ends of the third and fourth quarters of 2025, system-wide balances dropped by about $100 billion and rebounded the following day by about the same amount. On the same dates, repo investments by money market funds and other cash investors in the Fed's Overnight Reverse Repurchase Facility (ONNRP) jumped from near zero by about $50 billion and $100 billion, respectively, and immediately after the quarter ends fell by about the same respective amounts.[8]

When the quantity of reserves suddenly falls on quarter ends, some large U.S. dealer

[6]See Anderson and Na (2024); Banegas and Tase (2016); Logan and Schulhofer-Wohl (2025); Anbil, Anderson, and Eyal (2026).

[7]See Reserve Demand Elasticity, Federal Reserve Bank of New York. The February 2016 update of this series of reports includes the remarks (1) "The elasticity of the federal funds rate to reserve changes is very small and statistically indistinguishable from zero," and (2) "The estimate suggests that reserves remain abundant."

[8]Data: ONRRP from FRED. Estimated daily reserve balances are as reconstructed by Lou Crandall of Wrightson Associates. As at the end of October 2025, reductions in period-end reserve balances caused by FBO window dressing are not necessarily fully offset by increases in ONRRP balances. As explained to me by Lou Crandall, the Fed's quarterly financial statements generally show some statement-date increases in "other deposits" held by government-sponsored enterprises (GSEs) and in the Fed's FIMA repo facility.

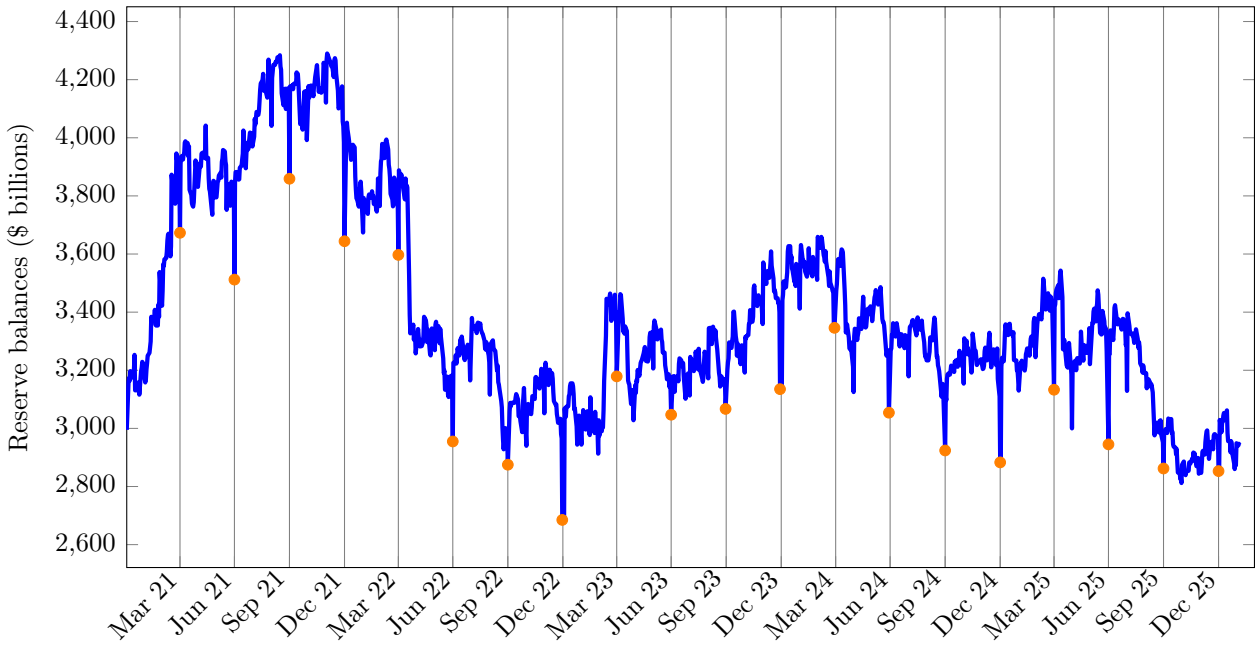


Figure 4: Total reserve balances, highlighting balances at quarter ends with orange dots. Data: Reconstructed daily total reserve balances, Louis Crandall, Wrightson Associates.

banks are reluctant to expand their balance sheets enough to fill the resulting shortfall in market funding with reserves obtained from the Fed's liquidity facilities, given the associated stigma and the cost to their shareholders of increasing their own required capital. Logan (2025) notes that these costs are higher than would be necessary if the Fed would centrally clear its repo operations, which allows dealer banks to net their Fed repos against their customer reverse repos.

Although some large U.S. bank dealers may have sufficient capital headroom to arbitrage period-end rate spikes with reserves obtained from the Fed's liquidity facilities, this is apparently not sufficiently enticing to be done at full scale, given the associated stigma. The quantity of repo funding demanded from the Fed with SRPs at the two most recent stressful period ends, October 31, 2025 and December 31, 2025, were about $50 billion and $75 billion, respectively. These are not small amounts, but much less than the quantity of Treasury repos conducted at rates well above the SRP rate.[9]

The demand-driven approach entails some tolerance by the Fed for volatility in market repo rates. The Fed's Manager of the System Open Market Account (SOMA), Roberto Perli, remarked that "Some repo rate volatility is not problematic and is arguably beneficial for allowing markets to send signals on market conditions. However, if repo funding costs become too volatile and unpredictable, the likelihood of forced liquidations of repo-financed Treasury positions increases, and that would create instability in the Treasury cash market

[9]Data: Federal Reserve Bank of New York, Repo Operations.

and related market segments."[10] Related to this, Deputy SOMA Manager Julie Remache[11] explained that

At lower levels of ample, we might observe increases in money market rates relative to administered rates. There may also be greater usage of SRP operations to provide liquidity, when it is desired by counterparties. However, drops in reserve supply below the ample region could result in elevated levels of volatility in money market rates and threaten the Fed's control of its policy rate. Excessive volatility can raise the risk of strained market functioning, making it harder for market participants to redistribute liquidity and access funding, including financing for Treasury securities.

Significant rate volatility has also sometimes generated commentary in the press that erodes perceptions of the Fed's ability to implement its monetary policy effectively. Perceptions matter and can play a role in policy.

2.3 The benefits and costs of a large balance sheet

With a conservative supply-driven approach, supported by Gagnon and Sack (2020), banks have plenty of incentive to arbitrage differences between IORB and market repo rates, leading to a smooth and tight relationship between the Fed's administered rates and virtually no risk of a costly liquidity crunch. In addition, with abundant reserves, banks avoid the costs of closely managing the sufficiency of their reserve balances. When answering a question at the February 2026 meeting of the National Association for Business Economics, Federal Reserve Governor Chris Waller applied a colorful metaphor, saying "You don't want banks ... digging around in the couch cushions, looking for money. This is massively inefficient and stupid."[12] An abundant supply of reserve balances also provides a liquidity cushion when stresses might otherwise lead banks to rely more heavily on short term credit that could suddenly disappear. That risk can also be addressed with liquidity regulations, but in practice the largest banks rely heavily on reserve balances to meet their liquidity requirements (Bush and others, 2019). Relying on liquidity regulations rather than abundant reserves has the disadvantage that Bank A may be reluctant to lend its reserve balances to Bank B when Bank B is undergoing a stressful liquidity event because Bank A wants to continue to meet its regulatory liquidity requirements. This issue is discussed in more detail in Section 6.

On the other hand, a large Fed balance sheet has been controversial (Warsh, 2025; Miran, 2025). Among the costs of abundant reserve balances, the Federal Open Market Committee

[10] See Perli (2025).

[11] See Remache (2026).

[12] See Amara Omeokwe, "Fed's Waller Says March Rate Call Depends on Labor Market," *Bloomberg*, February 23, 2026. See also Waller (2025).

(2018) noted that

"... Potential drawbacks of an abundant reserves regime included challenges in precisely determining the quantity of reserves necessary in such systems, the need to maintain relatively sizable quantities of reserves and holdings of securities, and relatively large ongoing interest expenses associated with the remuneration of reserves. Some noted that returning to a regime of limited excess reserves could demonstrate the Federal Reserve's ability to fully unwind the policies used to respond to the crisis and might thereby increase public acceptance or effectiveness of such policies in the future."

Waller (2025) explains how the interest expense of the Fed can be viewed, in effect, as a pass-through to banks from the Treasury, given that banks could equally well earn about the same interest by holding Treasury bills rather than reserves, at the about same expense to the government. Reserves offer better liquidity services to banks than T-bills. Nevertheless, the FOMC's minutes show the Committee's sensitivity to its interest expense. Transcripts of FOMC meetings also show that some members of the FOMC recognize the political costs of having a large balance sheet.[13] Berentsen, Marchesiani, and Waller (2014) find that when abundant reserve balances lead to political costs for a central bank, the central bank naturally retreats to a corridor approach that keeps the supply of reserve balances somewhat restrictive and offsets some of the central bank's interest expense paid at its deposit rate with interest earned at its higher lending rate.

There may also be a ratchet effect on the demand for reserves (Acharya, Chauhan, Rajan, and Steffen, 2022; Acharya and Rajan, 2022; Nelson, 2026b), by which ample amounts of reserve balances cause banks to become "addicted" to ever larger amounts of reserve balances. For example, abundant reserves reduces the incentives of banks to devote resources to systems and processes for quickly accessing reserve balances from the market or from the Fed's liquidity facilities. Banks may also become more dependent on their opening-of-day balances to process their daily outgoing payments, rather than developing methods that make more efficient use of incoming payments, as discussed in Sections 3 and 8. Equipped with abundant reserves, banks may provide their customers with demandable deposits, credit lines, and other liquidity intensive products that are costly to reverse in the short run, adding to the ratchet effect. So, even if the costs of a larger central bank balance sheet are proportionately small, the ratchet effect can lead the absolute size of the costs to eventually become large.

Duygan-Bump and Kahn (2026) summarize the tradeoffs: "A larger balance sheet stabilizes rate control and reduces the need for intervention but expands the central bank's structural footprint. A smaller balance sheet limits that footprint yet either requires more active liquidity management or tolerates greater rate variability."

2.4 Why has the demand for reserves gotten so large?

One of the objectives of this paper is to answer the commonly asked question: "What has changed so much since 2007, when total reserve balances were around $10 billion, that has caused the financial system to now need trillions of dollars of reserve balances?" The

[13] See, for example, the Transcript of the Meeting of the Federal Open Market Committee held from April 30 to May 1, 2019 (Federal Open Market Committee, 2019).

primary factors driving this massive increase in the minimum necessary quantity of reserves are:

1. A GSIB is now far more reluctant to draw intra-day liquidity from the Fed. Doing so could be viewed as a sign that it does not meet the self-reliance requirements of post-GFC intra-day liquidity regulations. For example, in the first two quarters of 2007, when total reserve balances were under $10 billion, total daylight overdrafts by the ten largest banks alone averaged $120 billion per day, mostly in the Fedwire Funds payment system (Badev and others, 2021). Since 2020, total peak daylight overdrafts in Fedwire Funds has averaged under $5 billion.[14] Pre-GFC, the vast majority of daylight overdrafts were conducted by the ten largest banks, whereas in recent years the amount of daylight overdrafts by the ten largest banks is typically less than that of other banks (Badev and others, 2021). I will use some illustrative theory from Duffie, Singh, and Wang (2026) to show that, in the absence of liquidity savings mechanisms, the amount of reserves required for timely payments can be lowered significantly if daylight overdrafts are not viewed by banks as costly. The resulting increase in daylight overdrafts, as modeled, is far smaller than the associated reduction in required opening balances. Daylight overdrafts are normally secured by collateral and have no direct fees.

2. Pre-GFC, large banks worked hard to lend the reserves that they don't need to manage their payments, given the alternative of receiving no interest for them from the Fed (Hamilton, 1996; Ashcraft and Duffie, 2007; Judson and Klee, 2009).[15] Because the Fed now remunerates banks for their reserves at roughly market interest rates, regardless of the amount of reserves they hold, there is a far smaller incentive for banks to lend their excess reserves. Indeed, reserves are an especially attractive investment opportunity for some of the largest banks when the yield curve is downward sloping (Chien and Stewart, 2024). For exmple, with recent reductions in IORB and increases in the slope of the term structure of interest rates, the quantity of reserve balances held by J.P. Morgan Chase has declined dramatically, far more than proportionate to the contemporaneous reduction in the total suppy of reserve balances. According to reporting in *Financial Times* in late 2025, "JPMorgan Chase has withdrawn almost $350bn in cash from its account at the Federal Reserve since 2023 and ploughed much of it into US government debt, as the bank tries to defend itself against rate cuts that threaten to erode its profits. JPMorgan, which has more than $4tn in assets, slashed

[14] See Federal Reserve Board data and Badev and others (2021). There was one exceptional day in March 2020 when the total daylight overdrafts of the ten largest banks was significant, approximately $100 billion (Badev and others, 2021).

[15] For example, Judson and Klee (2009) find that "DIs [depository institutions] typically hold excess balances on a precautionary basis in light of uncertain flows in and out of their reserve accounts. ... Excess reserve balances are held voluntarily by DIs, typically as a cushion to guard against overdrafts. Excess balances are calculated as the difference between balances held at the Federal Reserve and ROB [required opening balances]. Holding balances in excess of required operating balances helps DIs to avoid overdrafts but carries opportunity costs. In contrast, closely managing excess reserve balances in order to minimize excess holdings while still avoiding overdrafts entails costs associated with staffing reserves management activities. As a general proposition, for larger banks, the benefits of managing excess balances closely frequently outweigh the costs, while for smaller institutions the costs generally appear to outweigh the benefits."

its balance at the Fed from $409bn at the end of 2023 to just $63bn in the third quarter of this year, according to data compiled by industry data tracker BankRegData."[16] An implication is that a significant component of the initial $409 billion position in reserve balances was treated by JPMorgan as part of its investment portfolio, as opposed to balances needed for operations.

3. On top of the just-mentioned post-GFC loss of interest-rate incentives for banks to reallocate their excess reserves, the frictional costs for interbank lending and borrowing have risen sharply since the GFC. The largest U.S. banks now face FDIC insurance fees of around 0.5% of their total liabilities.[17] and have higher capital requirements, including leverage-based capital requirements (Banegas and Tase, 2016).

2.5 Policies that may reduce the demand for reserves

Whether as a complement to the supply-driven or demand-driven approach to ample reserves, the Fed could reduce the required ample quantity of reserves with some combination of (1) temporary open market operations (TOMOs) that counteract unintended shocks to the supply of reserve balances, including those associated with period ends and changes in the size of the TGA; (2) revision of intraday liquidity regulations and their supervision; (3) adding a liquidity savings mechanism (LSM) to the Fed's largest payment system, the Fedwire Funds Service; and (4) tiering the interest rate paid by the Fed to banks on their reserve balances, so that banks prefer to lend reserves that they don't need for processing payments. As I will explain, the Fed has already engaged somewhat with adjustments to its liquidity regulations. On the other hand, liquidity regulations have a prudential purpose (Kashyap, Tsomocos, and Vardoulakis, 2024). It's not yet clear how much progress can be made on this front while meeting the financial-stability objectives of liquidity regulations. The Fed could also benefit from an in-depth cost-benefit analysis of the other three approaches. An LSM and tiering the remuneration of reserve balances would not be simple to communicate and implement operationally.

3 The interbank payment system

Modern bank-railed payment systems, at their core, are based on deposit accounts held by commercial banks at their central bank. This connects the customers of different commercial banks to each other, as illustrated in Figure 5. On an average day in the last quarter of 2025, roughly $7.3 trillion of payments of reserve balances were made using the Fed's two largest Fed payment systems, Fedwire and Fedwire Securities. By comparison, about 2 trillion euros per day are paid in the analogous Eurosystem large value payment systems.[18] The Fed's

[16] See "JPMorgan pulls $350bn from Federal Reserve to buy up Treasuries," Joshua Franklin, *Financial Times*, December 16, 2025.

[17] FDIC insurance fees currently range up to 42 basis points for the largest GSIBs, on top of special fee assessment of up to 13 basis points that followed the failure of SVB and other banks in March, 2023.

[18] Fedwire payments averaged about $4.5 trillion a day. Fedwire Securities is a system for transferring government securities, often in exchange for reserve balances (Copeland and Wang, 2024). The transfers of value reported by Fedwire Securities are quantities of reserve balances paid. Source for Eurosystem payment

other payment systems, including FedNow, are smaller and play a correspondingly small role in determining the size of the Fed's balance sheet.[19]

Figure 6 illustrates how a large payment relies on a transfer of reserve balances between banks. Alpha, a corporate client of Bank A, is making a payment of \$8 billion to another firm, Beta, a client of Bank B. Bank A instructs Fedwire to reduce the reserve balance of Bank A by \$8 billion and increase the reserve balance of Bank B by \$8 billion. After Bank B receives a payment message with the necessary details, it increases Beta's deposit balance at Bank B by \$8 billion.

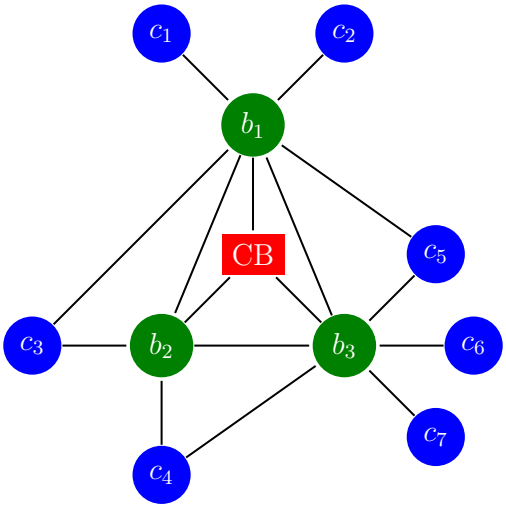


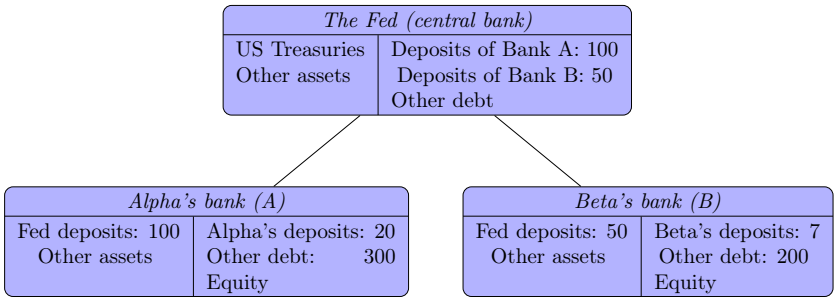
Figure 5: Schematic of a two-tiered bank-railed payment network. The customers of commercial banks are shown in blue. Commercial banks are shown in green. The central bank is labeled "CB." Payments from a customer of one bank to a customer of another bank are made via a sequence of debits and credits of commercial bank and central bank deposit accounts along the path between the ultimate payer and receiver.

The Fed processes payments in Fedwire as soon as they are requested by banks, an approach known as real time gross settlement (RTGS). RTGS payment systems such as Fedwire, Fedwire Securities, and FedNow contrast with deferred net settlement payment systems such as the Automated Clearing House (ACH), which delay payments so that they can be batched. Batching allows a bank to pay only the net of its outgoing payments over its incoming payments during the batching period, thus reducing the amount of central-bank deposit balances that a bank needs at the beginning of a day to cover its payment obligations on that day. On the other hand, deferred net settlement involves delays and a risk that the payment might not ever be made because the payer defaults before the batching time. RTGS avoids these costly delays, but requires larger initial central bank deposit balances.

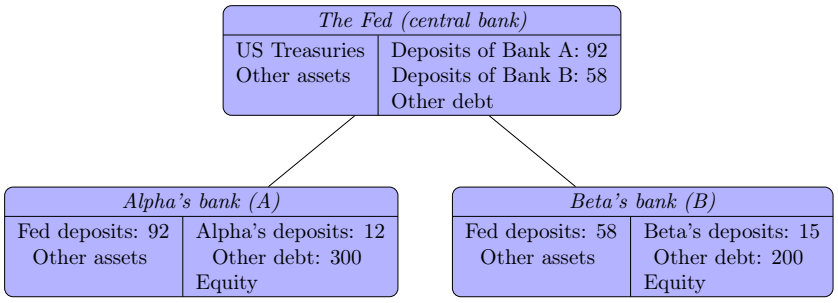
Many central banks use liquidity savings mechanisms (LSMs) that reduce the amounts

statistics: European Central Bank.

[19]In addition to Fedwire, Fedwire Securities, and FedNow, the electronic payment systems operated by the Fed include FedACH® (Automated Clearinghouse), National Settlement Service (NSS), and Check Clearing Services (Federal Reserve Bank of New York). Leistra (2026) explains U.S. the architectures used in U.S. payment systems and discusses the costs and benefits of some alternative proposed approaches.



(a) Central bank and commercial bank balance sheets before the payment by Alpha to Beta. Alpha's bank (A) has, among its assets, $100 billion in reserve balances at the Fed. The liabilities of Bank A include Alpha's deposit of $20 billion. Beta's Bank (B) is analogous to Bank A. The Fed has assets that include U.S. Treasury securities. Among the Fed's liabilities are the reserve balances of banks A and B.



(b) Balance sheets after Alpha's payment to Beta. With the payment of $8 billion by Alpha to Beta, Alpha has $12 billion remaining in deposits. Bank A, having reduced its reserve balances at the Fed by $8 billion, has $92 billion remaining in reserve balances. The reserve balances of Bank B at the Fed have increased by $8 billion to $58 billion. Beta's deposit account balance at Bank B has increased by $8 billion. The accounting equity of each bank is unaffected by these payments because of the offsetting changes in assets and liabilities.

Figure 6: How a payment is made using an RTGS payment system such as Fedwire.

of central bank deposits needed to process a given volume of payments (World Bank Payment Systems Development Group, 2020). LSMs include incentives for banks to place large payments in a queue in which they can be offset against incoming payments, reducing the initial quantity of balances needed to process the payments. Section 8 explains how this works. The large-value payment systems of the Eurosystem, United Kingdom, Japan, and Canada use LSMs.[20]

Some payment system operators also have throughput rules that require each bank to process at least a stipulated fraction of its daily outgoing payments by a given time of day (Buckle and Campbell, 2003). Rather than a formal throughput rule, Canada's large value payment system, Lynx, monitors compliance with throughput "targets" (Bank of Canada, 2022).

A consortium of about 40 of the largest U.S. banks have access to CHIPS, a private-sector large value payment system operated by The Clearing House. CHIPS handles roughly $2 trillion of payments per day, of which 95% are international. By exploiting both netting

[20]See, for example, Bank of England (2021), Bank of Canada (2022), and Annex 1 of European Central Bank and Bank of Japan (2017).

and a liquidity savings mechanism, CHIPS achieves high payment efficiency (The Clearing House, 2025).

The Fed's largest payment systems, Fedwire and Fedwire Securities, do not use liquidity savings mechanisms or throughput rules.[21] The Fed does sponsor a forum in which banks discuss how to improve the timeliness of their payments to each other (Payments Risk Committee, 2019). Section 8 explains the potential for LSMs to reduce the quantity of reserve balances needed by the payment system. This approach seems likely to improve the effectiveness of monetary policy implementation when reserves are not abundant (Buckle and Campbell, 2003).

Fedwire Securities handles an extremely large amount of payments of reserve balances in exchange for securities, about $2.6 trillion per day in the third quarter of 2025. However, relatively little timing discretion applies to these payments, so there is less scope for LSM benefits. Most Fedwire Securities payments are made early in the day, as shown in Figure 7a from Copeland and Wang (2024). For example, the majority of the volume of repurchase agreements (repos) and Treasury auction settlements are processed early each day.

With Fedwire, as opposed to Fedwire Securities, payment times are widely distributed across the day, as shown in Figure 7b, and large banks time their payments strategically. For example, the quantity of outgoing Fedwire payments made by a bank at a given time of day is sensitive to recently received incoming payments (Afonso, Duffie, Rigon, and Shin, 2022), as shown in Figure 8. When large banks fear that they may not have enough central bank deposits to process their outgoing payments quickly while meeting their intraday regulatory liquidity requirements, they can resort to slowing down ("throttling") their payments to other banks (Bech and Garratt, 2003a,b). This can cause feedback by which some banks wait to receive enough payments from other banks before processing much more of their own outgoing payments, potentially leading to gridlock (McAndrews and Potter, 2002).

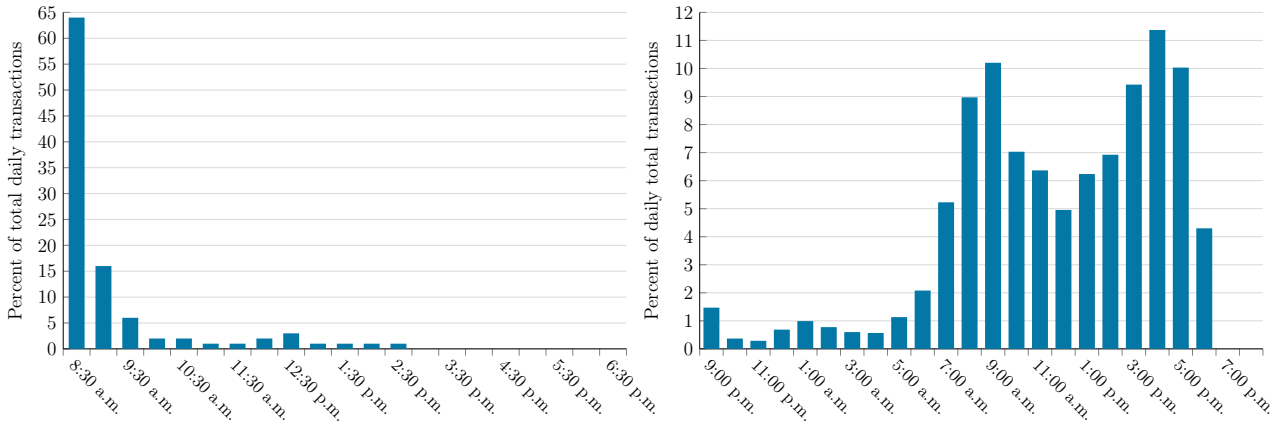
As evidence of this "receipt-reactive" payment timing, Figure 8 shows that a large bank's incoming amount of Fedwire payments in the last 15 minutes predicts its outgoing amount of Fedwire payments in the next minute, based on Tobit estimates of

$$\log\left(1+P_{imt}\right)=\beta_0+\beta_1\log(1+X_{it})+\gamma_i+\gamma_t+\sum_j\gamma_{mj}^j+u_{imt}, \tag{1}$$

where P_{imt} is the total dollar value of payments from bank i to its counterparties in minute m on day t; X_{it} is the total amount of the incoming payments to bank i during the previous 15 minutes; γ_i is a sender-bank fixed effects; γ_t is a date fixed effect; γ_m^j is, for each j, a fixed effect that is based on whether m is in the j-th specified interval of time during the day; and u_{imt} is an unexplained residual.[22] As shown in Figure 8, when the total reserve balances of large banks are lower, the estimated dependence of outgoing payments on incoming payments is significantly higher.

[21] The Fed does assess a late-payment surcharge for Fedwire payment requests received after 5pm Eastern, but this fee is currently only $0.26 per message.

[22] Of the time-of-day fixed effects, γ_m^{open} is an indicator variable equal to 1 if m is between 9:00 pm and 9:30 pm ET on the preceding calendar day, γ_m^{early} is equal to 1 if m between 9:30 pm ET on the preceding calendar day and 6:00 am ET, γ_m^{eod} is equal to 1 if m is between 6:00 pm and 6:30 pm ET, and there are similar fixed effects for each 30-minute interval between 2:00 pm and 6:00 pm ET. Standard errors are adjusted for heteroskedasticity and clustered at the bank (sender) level.



(a) Distribution of payment times in Fedwire Securities.

(b) Distribution of payment times in Fedwire Funds Service.

Figure 7: Distribution of payment times in Fedwire Seurities (a) and Fedwire Funds Service (b). The chart shows the percent of all daily transactions occurring in each 1.5-hour period over the operational days of the respective systems. The sample period is the second quarter of 2024. Figure source: Redrawn from Copeland and Wang (2024).

As total reserve balances decline and the risk of payment delays rises, banks and their affiliated large dealers become more reluctant to provide funding to others in money markets, leading occasionally to unpredictable sudden increases in market interest rates, as happened during the GFC, in mid-September 2019, and at other times.[23]

Copeland, Duffie, and Yang (2025a) analyze the connection between Treasury repo rate spreads and the timing of payments to the 10 bank holding companies that are most active in the Treasury repo market, which they call "the dealer banks." They find that when the dealer banks have relatively low opening-of-day balances and have recently been receiving their large incoming payments much later than normal, they are much more likely to offer funding to counterparties at elevated repo rates. Yang (2020) provides a supporting theory. This effect is illustrated in Figures 9 and 10.

Figure 9 shows in red (left-hand axis) the spread of SOFR to IORB, and in blue (right-hand axis) the total reserve balances of the ten dealer banks. As shown, when the dealer-bank total reserve balances are low, repo rates are elevated and can spike well above IORB. This relationship between dealer-bank balances and SOFR-IORB spreads is a general property of the 2015-2025 sample analyzed by Copeland, Duffie, and Yang (2025a) using models based on least-squares prediction of repo-rate spreads, quantile regression prediction of tail repo-rate spreads, and logit estimation of the likelihood of spikes in repo-rate spreads. A key predictor of elevated repo rate spreads is the extent of delays in incoming payments to the ten dealer banks.

On September 17, 2019, with relatively little warning in terms of repo-rate spreads, SOFR jumped to 315 basis points above IORB. In March 2020, the COVID shock also caused a sudden jump in SOFR-IORB. However, the COVID shock also caused the Fed to purchase massive amounts of Treasury securities to mitigate dysfunction in Treasury markets, a by-product was the creation of a correspondingly massive amount of reserve balances. As shown

[23]See Ashcraft, McAndrews, and Skeie (2011); Afonso, Kovner, and Schoar (2011); Copeland, Duffie, and Yang (2025a); Yang (2020).

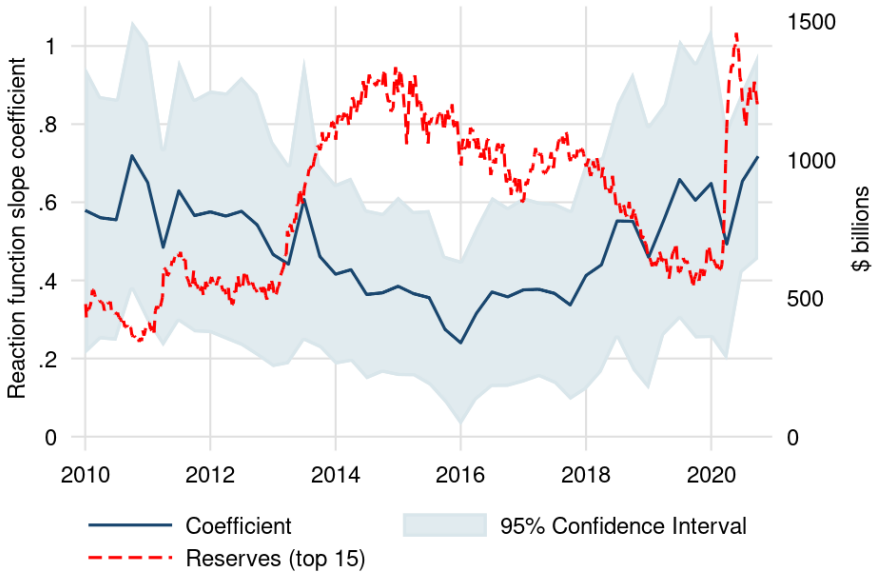


Figure 8: Quarterly Tobit coefficient estimates of the reaction-function coefficient $\hat{\beta}_1$ in Equation 1. The data are for the largest 15 Fedwire payment entities, as defined by average daily total dollar value of payments sent (dashed red line), 2010-2020. Source: Afonso, Duffie, Rigon, and Shin (2022). Data: Internal Federal Reserve accounting records and Fedwire Funds Services.

in Figure 9, the resulting abundance of reserve balances then caused SOFR to smoothly match IORB.

As shown in Figure 10, the ten dealer banks have experienced substantial variation over the past decade in the time of day by which they have received the first half of their incoming daily payments of reserve balances, a standard metric for payment timing.[24] From peak to trough, this payment timing measure has varied by roughly 250 minutes between the beginning of 2015 and the end of February 2026. The largest delay in this ten-year sample occurred on September 17, 2019, a day on which system-wide reserve balances reached its sample-record low and on which, as shown in Figure 9, SOFR jumped to a record-high spread to IORB. In the interdealer market that day, repo rates soared by roughly 1000 basis points above IORB. As shown in Figure 10, the next largest payment delay in the sample occurred at the onset of the COVID pandemic, on March 17, 2020, when SOFR spiked above IORB by 44 basis points. As shown, this measure of payment timing delays rises systematically during each of the Fed's two quantitative tightening periods, September 2017 to September 17, 2019, and June 2022 to December 2025. That said, the total quantity of reserve balances did not decline throughout the second QT period, mainly because up to $2.5 trillion of reserve balances were replaced with ONRRP balances.[25]

Figure 11, from Copeland, Duffie, and Yang (2025a) illustrates the mechanism at work. The vertical axis shows the payment-delay metric from Figure 10 for the ten dealer banks, restricted to payments received from the next 90 largest banks in the system as measured by sample-average reserve balances. The horizontal axis corresponds to the total beginning-

[24]See Armantier, McAndrews, and Arnold (2008), McAndrews and Kroeger (2016), and and Copeland, Molloy, and Tarascina (2019).

[25]Data: FRED.

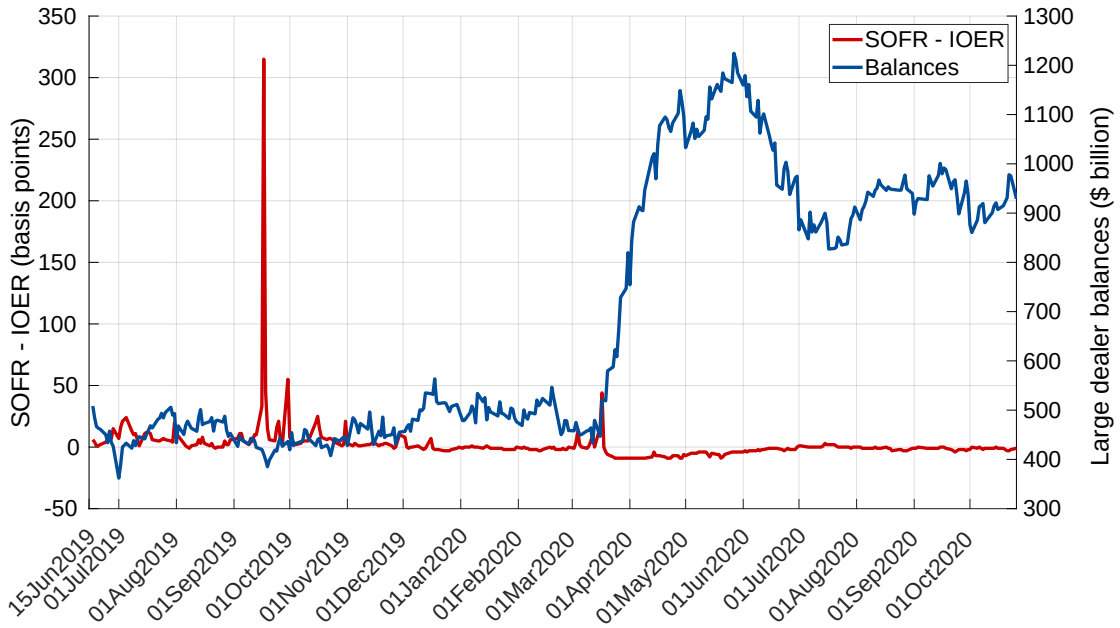


Figure 9: Reserve balances and the spread of SOFR over IORB, known at the time as "IOR." . The reserve balances of the ten dealer banks are shown in blue (right axis). The spread of SOFR over IORB is shown in red (left axis). Figure source: Copeland, Duffie, and Yang (2025a). Data: Fedwire Funds Service, FRBNY.

of-day balances of these next-90-biggest banks. Each point in the scatter plot corresponds to one of the days in the sample. The least-squares fitted line shows a clear and significant reduction in delays of payments to the dealer banks as the balances of the other large banks grow larger. The points colored in red correspond to days on which SOFR-IORB attained its highest 20 levels in the sample. The top left observation (in red) is for September 17, 2019. Naturally, when system-wide balances decline, banks delay their outgoing payments. Large funding-market intermediaries then perceive a higher shadow cost for meeting the constraint of having sufficient balances. This effect is also manifested in other funding markets (Correa, Du, and Liao, 2025).

4 Searching for "ampleness"

The Fed's framework for implementing its monetary policy has changed significantly over the past two decades. Before the GFC, the Fed did not pay interest to banks on their reserve balances. The Fed was able to steer market interest rates by keeping the total supply of reserve balances low enough that many banks had scarcely enough reserves to meet their reserve requirements and intra-day payment needs. The Fed found this to be a challenging approach to steering interest rates (Hamilton, 1996; Ashcraft and Duffie, 2007). Remarkably, in the first half of 2007, total reserve balances were typically under $10 billion (FRED).

Pre-GFC, whenever bank i had a shadow price λ_i for reserve balances above the shadow price λ_j of some other bank j, bank i had an incentive to borrow reserves from bank j at an interest rate r_{ij} somewhere between λ_i and λ_j. The Fed adjusted the supply of reserve

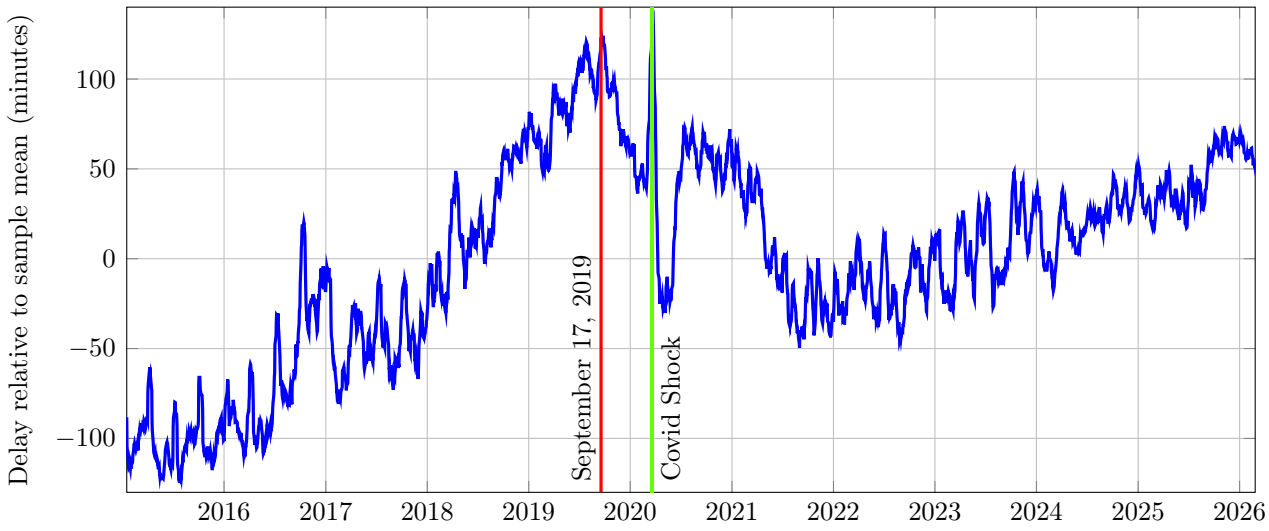


Figure 10: The 10-day trailing average of intra-day payment delays to the repo dealer banks. The lagged average of the time of day by which the ten largest repo-active bank holding companies had received half of their large ($\geq$ \$100 million) incoming payments, relative to sample mean, in minutes. The observation shown for date t is the average delay for business days $t-10$ through $t-2$ (inclusive). The first peak is at September 16, 2019, the beginning of the three-day repo-market liquidity crunch of September 2019. The second peak is March 12, 2020, corresponding to the Covid shock. OLS regression: A one-standard-deviation increase in the trailing average payment delay predicts a 7-basis-point increase in SOFR−IOR. Source: Work in progress by Copeland, Duffie, Wang, and Yang (2025). Data: Fedwire Funds Service.

balances so that the average of the rates executed in the federal funds market each day was as close as possible to the Fed's target rate. This caused the Fed to maintain a low total amount of reserves. Banks made relatively liberal use of daylight overdrafts of their Federal Reserve deposit accounts to meet their intraday payment needs (Badev and others, 2021).

The Fed responded to the GFC with quantitative easing (QE), paying for the large quantity of securities that it purchased with new reserve balances. A side effect of QE was thus "abundant" reserves, meaning that, on average, banks had far more reserves than they needed to process their payments. Additional rounds of QE eventually pushed total reserve balances to a peak of \$4.2 trillion in September 2021, about 400 times pre-2008 levels.

After QE made reserves abundant, shadow prices for reserve-balances sufficiency fell close to zero most of the time. In order to steer market rates, the Fed was therefore forced to remunerate balances at an interest rate (IORB) that anchors the opportunity cost to banks for lending. Absent frictions, an abundant supply of reserves and competition between banks would keep the risk-free overnight market rates close to IORB.[26] Frictions do, however, interfere with this simple "floor system" for monetary policy implementation.

Among these frictions are leverage-based capital requirements. When reserves are more abundant, GSIBs intermediating the repo market have more incentive to lower their capital

[26]Even before the GFC, the Fed found it difficult to steer rates without paying interest on reserve balances. The Fed obtained the authority from Congress to do so in the 2006 Financial Services Regulatory Relief Act. This Act authorized the Fed to pay interest only beginning in 2011, but the authorization date was moved up to October 2008 when GFC-induced QE created abundant reserves.

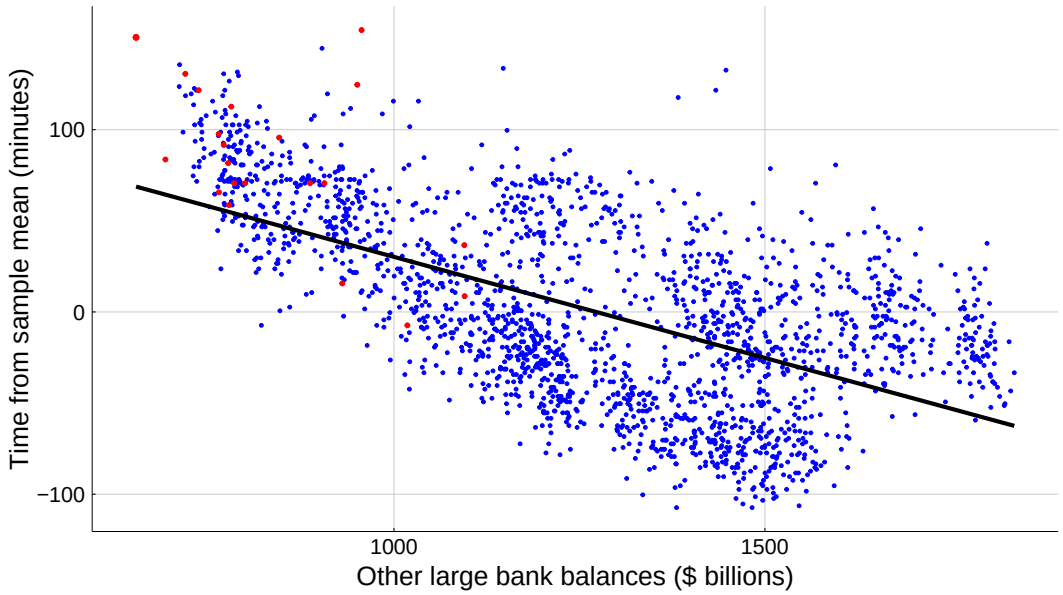


Figure 11: Payment delays. Time, relative to sample average, at which half of payments to the ten largest dealer banks is received from the next 90 largest banks, as measured by sample average reserve balances, regressed on opening balances of next 90 largest banks. $R^2 = 0.36$. The slope coefficient, -0.11, is estimated with a standard error of 0.00325. Red dots are days on which SOFR-IOR attained its highest 20 levels. The top left observation (in red) is for September 17, 2019. Figure source: Copeland, Duffie, and Yang (2025a). Data source: Fedwire Funds Service.

requirements by reducing their stocks of reserve balances. They are willing to intermediate the market by accepting investments of reserves from money-market mutual funds only at repo rates that are sufficiently below IORB. The Enhanced Supplementary Leverage Ratio (eSLR) rule and the Tier 1 leverage rule impose the same proportional capital requirement for reserves as for other assets.[27] For GSIBs sensitive to leverage-based capital requirements, holding large amounts of reserve balances generates a non-trivial opportunity cost for bank shareholders (Duffie, 2022). To prevent the large quantity of balances created by QE from driving market rates significantly below IORB, the Fed introduced the Overnight Reverse Repurchase Facility (ONRRP), where money-market funds (among other non-bank investors) are able to invest cash at an overnight interest rate administered by the Fed. When banks hold so much reserves that market rates fall too far below IORB, the Fed increases the ONRRP rate, leading money funds to increase their cash investment at the ONRRP and reduce their lending to banks and securities dealers. This shift to the ONRRP extinguishes reserve balances, dollar for dollar. The ONRRP rate is thus a "subfloor" rate that helps the Fed steer market rates when reserves are abundant.[28]

The focus of this paper, however, is the opposite situation in which reserve balances threaten to be insufficient, facing some banks with a substantial shadow price for reserve

[27] As the Fed announced in November 2025, U.S. banking regulators have significantly lowered the eSLR capital requirement.

[28] The fact that certain federal agencies, particularly Federal Home Loan Banks, have reserve balances that are not remunerated by the Fed also puts downward pressure on market rates.

balances and causing market rates to become volatile and significantly *above* IORB. Avoiding this outcome places a floor on the required total quantity of reserve balances and thus on the size of the Fed's balance sheet.

The FOMC has adopted an "ample-reserves" approach, aiming to reduce balances to the minimum level consistent with market rates being driven primarily by IORB.[29] Initially, the Fed had in mind a notion of ample reserves that would be achieved when the slope of the demand function mapping balances to market rates transitions from virtually flat (abundance) to moderately downward sloping (Afonso, Giannone, Spada, and Williams, 2025). This demand function is known as the Poole Curve (Poole, 1968). For example, Figure 12 shows a conceptual version of the Poole curve, in green, whose slope is increasingly negative as the total quantity of reserve balances drops below $1.4 trillion dollars, the lowest level achieved during the Fed's balance-sheet "normalization" over the period August 2017 to September 2019.

Ideally, as total reserve balances dropped below $1.4 trillion in September 2019, the increasing upward movement of SOFR−IORB along the green curve would have signaled to the FOMC that reserves were no longer abundant, and that an ample level of reserve balances had been achieved. At that point, the FOMC would have stopped reducing the size of its balance sheet. In actuality, when total reserve balances reached $1.4 trillion, there was a liquidity crunch. On September 17, 2019, SOFR−IORB spiked to 315 basis points. During the day, interdealer repo rates jumped by roughly 1,000 basis points. The corresponding sudden shift to a different conceptual Poole Curve is shown in the blue segment of Figure 12, corresponding to total reserve balances below $1.4 trillion. Large dealer banks apparently joined in a self-fullfilling equilibrium expectation that their incoming payments would be further delayed and, as a consequence, there was a corresponding jump in the shadow cost of meeting the constraint of having sufficient reserves. Yang (2020) provides a supporting theory. The Fed reacted to the spike in repo rates on September 17, 2019, by ensuring that banks had an additional $75 billion of reserve balances by October 7, 2019.

In short, theory and evidence suggest that slope of the Poole curve need not adjust gradually enough to rely on its slope as a signal of the onset of insufficient reserve balances.

Figure 13, from Duygan-Bump and Kahn (2026), further illustrates this difficulty. The left panel shows how the estimated sensitivity of Treasury repo rates (spread to the ON-RRP rate) depends on the quantity of reserves (normalized by the quantity of Treasuries outstanding, which scales for the size of a key related market in which reserves are used as the cash medium of payment). As shown, this sensitivity can change abruptly. The fact that there exist many days on which the sensitivity was high suggests that reducing reserves until the slope reacts moderately has actually resulted in some overshooting. The right panel of Figure 13 shows that when this overshooting happens, the volatility of the repo rate spread also climbs sharply. The general picture is one in which the right tail of repo rate distortions is very sensitive to reserve balances, especially the quantity of balances held by the largest repo dealers (Copeland, Duffie, and Yang, 2025a; Correa, Du, and Liao, 2025).

[29] In its press release on January 2019, the Federal Reserve Board stated: "The Committee intends to continue to implement monetary policy in a regime in which an ample supply of reserves ensures that control over the level of the federal funds rate and other short-term interest rates is exercised primarily through the setting of the Federal Reserve's administered rates, and in which active management of the supply of reserves is not required."

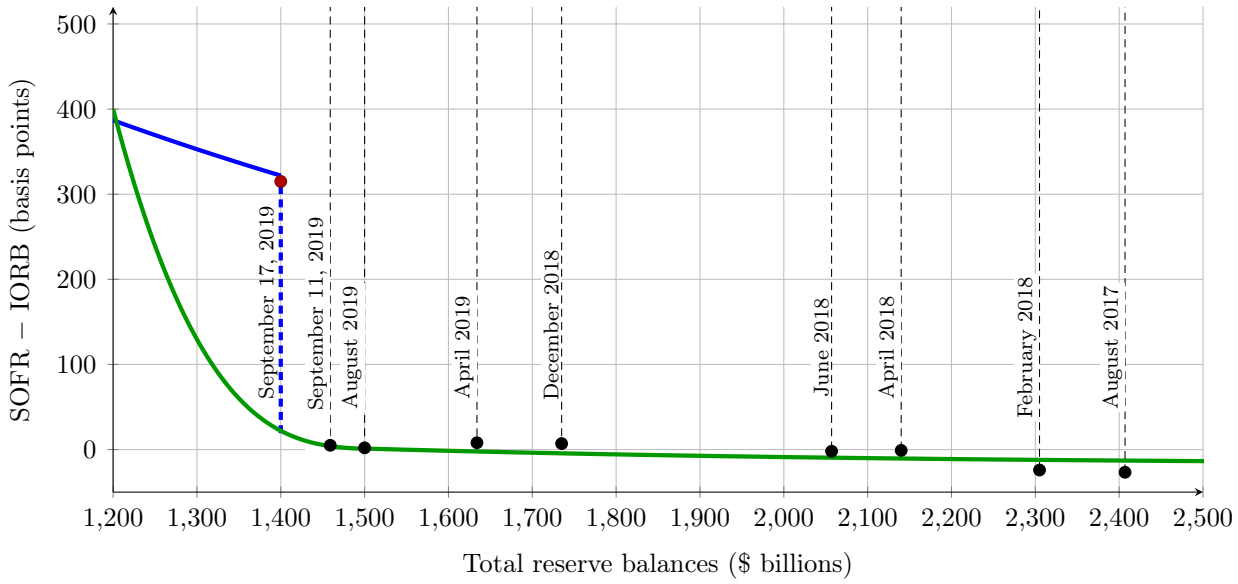


Figure 12: Two conceptual versions of the Poole curve, showing the relationship between market interest rates and reserve balances during the balance-sheet normalization conducted by the Fed over the period from August 2017 to September 17, 2019. For dates before September 2019, the red dots correspond to the actual monthly average repo rate spread, SOFR−IORB, and the actual monthly level of reserve balances. For September 11 and 17, 2019, the red dot corresponds to the data for those dates. SOFR−IORB was 5 basis points on September 11, 2019 and 315 basis points on September 17, 2019. The green version of the Poole curve shows a continuous relationship between the level of reserve balances and the slope of the curve. Conceptually, the Poole curve with the blue segment at levels of balances below $1.4 trillion corresponds to a self-fulfilling expectation of an equilibrium jump in SOFR−IORB when balances drop below $1.4 trillion. Data: Federal Reserve Bank of New York (SOFR), Federal Reserve Board, H.4.1 (reserve balances) and FRED (IORB).

Distinguishing between the Fed's notion of ample reserves before and after the liquidity crunch of September 2019, the Fed's former Vice Chair for Banking Supervision, Randall Quarles, remarked:[30]

"So, the Fed then had been following the view of, 'The balance sheet needs to be as small as possible, but not smaller. And we don't know exactly what that size is. But we'll see it when we kind of reach a kink in the demand curve as we begin to shrink reserves and shrink the size of the balance sheet. At some point, we'll see kind of a developing of a slope in the demand curve, and we'll know that we're there. And then we'll kind of expand it back out a little bit.' The disruption in September of 2019, as we were shrinking the balance sheet, was significant enough that I think the general view on the FOMC became, 'We don't want to ever get that close again' and movement away from the view of 'Let's shrink it until we're absolutely as close to that kink as possible.' It was, 'Well, let's just make sure we've got plenty of cushion before that ever happens again."

In summary, it has proven to be difficult for the Fed to predict the transition point between abundant and ample reserves. The FOMC has updated its implementation of the

[30]See Beckworth and Quarles (Beckworth and Quarles).

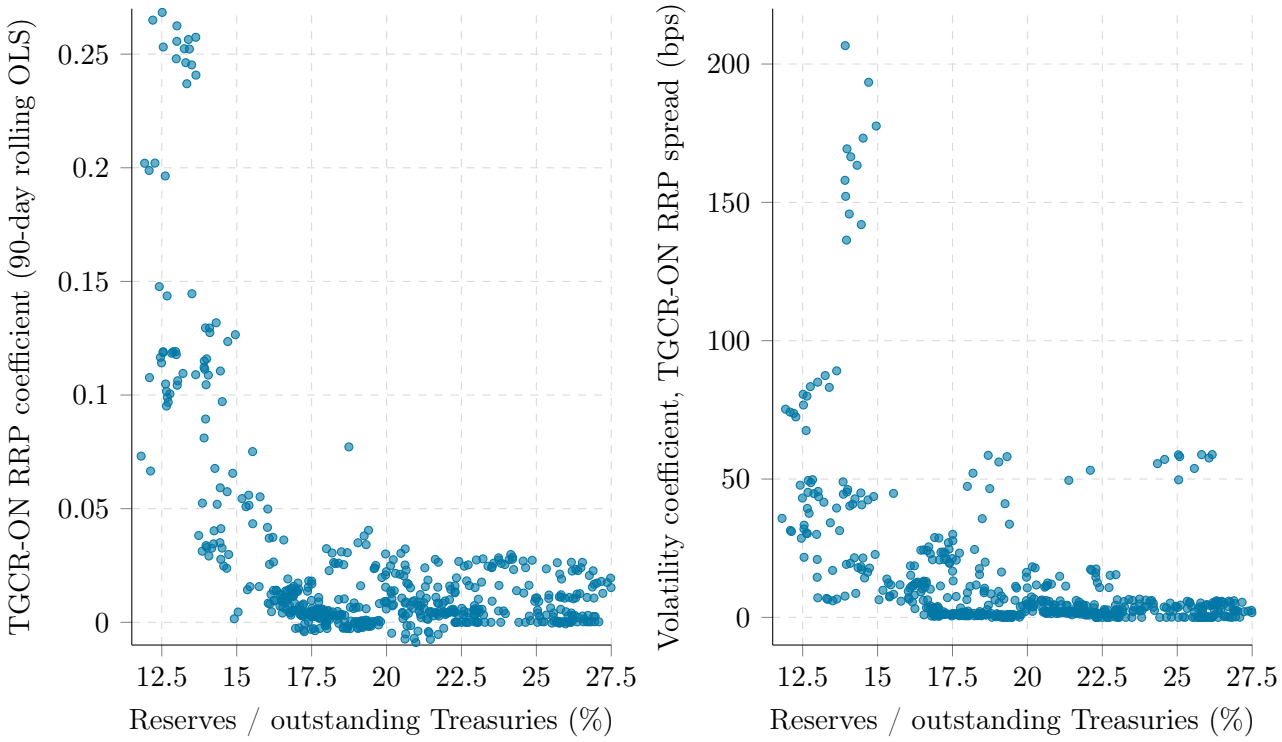


Figure 13: The OLS coefficient describing the estimated prediction of the Treasury repo rate (left chart) and repo rate volatility (right chart), on the ratio of reserve balances to the quantity of Treasuries outstanding. Both series use spreads of the Triparty General Collateral Repo Rate (TGCR) over the administered interest rate on the overnight reverse repurchase facility (ON-RRP). Data for the week of September 17, 2019 and the month of March 2020 are dropped. Volatility is the 90-day rolling standard deviation of the spread, sensitivity is a rolling OLS regression of the change in the spread on the change in the Treasury General Account (TGA), using the same methodology as Gissler, Hempel, Kahn, McCabe, and Narajabad (2025). Figure source: Redrawn from Duygan-Bump and Kahn (2026). Data sources: Federal Reserve Bank of New York Reference Rates, H.4.1. Factors Affecting Reserves, TreasuryDirect.

ample-reserves approach several times since 2017 (Board of Governors of the Federal Reseve System, 2019).

After the liquidity crunch of September 2019, the Fed introduced the Standing Repo Facility (SRF), at which banks and primary dealers can obtain funding from the Fed at interest rates slightly above IORB. Like the Discount Window, this ideally puts a ceiling on market rates, given the potential for banks and primary dealers to arbitrage by lending at rates above the SRF rate with funding obtained from the SRF. However, the SRF has had a stigma, much like that of the Discount Window, which discourages its use and thus limits its effectiveness at placing a ceiling on market rates. Even when market repo rates have jumped significantly above the SRF rate, the quantity of reserves drawn from the SRF has been modest.

For example, as shown in Figure 1, on October 31, 2025 Treasury repo rates were well above the SRF rate, which itself was 10 basis points above IORB. Despite this opportunity for bank dealers to obtain funding at the SRF rate and provide funding to counterparties

at much higher rates, a mere $50 billion of funding was obtained from the SRF.[31] Section 6 ascribes much of the reluctance of dealer banks to obtain funding from the SRF to stigma. At a banking industry symposium in November 2025, held under Chatham House rules, GSIB executives spoke of their reluctance to access SRPs for reasons other existential need.[32] This stigma is related in part to supervision of the Fed's liquidity regulations. For large bank dealers, capital regulations also inhibit use of the SRF. Because the Fed does not centrally clear its repo operations, bank dealers are unable to net their SRF repos against their customer reverse repos (Logan, 2025).)

In an effort to provide liquidity from the SRF more effectively when it is needed, in June 2025 the Fed introduced an additional morning SRF operation on each business day.[33] Then, on December 10, 2025 the Fed rebranded the SRF as Standing Repo Operations (SRP) and made this a full-allotment program, meaning that the Fed would no longer limit the aggregate amount of SRPs on a given day, although with some limits on the maximum amount of liquidity supplied to a given firm.[34]

As in the United States, quantitative tightening in the United Kingdom has significantly reduced the supply of reserves, from a peak of nearly one trillion pounds at the end of 2021 to 646 billion pounds at the end of January 2026.[35] Predictably, the reduction in sterling reserves has led to higher and more volatile gilt repo rates (Dolan and Roberts-Sklar, 2025). However, elevated repo rates have triggered a surge of demand for funding from the Bank of England's Short Term Repo (STR) operations (Ritchie and Hirai, 2026). The quantity of liquidity drawn from the STR is proportionately far more — even in absolute terms far more — than the quantities of liquidity obtained with the Fed's SRPs. The total supply of sterling reserves, including draws from the STR and other liquidity facilities, has nevertheless declined. Like the latest version of the Fed's Standing Repo Facility (SRP), the Bank of England's STR is provided on a full-allotment basis at an administered rate. Unlike the Fed's SRPs, which are overnight and priced at 10 basis points above IORB, the Bank of England's STR offers one-week repos at Bank Rate, the analogue of IORB.

A successful demand-driven approach to reserves ampleness seems close to achievable for the Fed. For example, the quantity of SRPs conducted on the last day of 2025, $75 billion, is a basis for optimism. It would be tempting to allow greater repo rate volatility in order to get dealers to access SRP funding more frequently and in higher volumes. By normalizing the use of SRP funding, the associated stigma would fall over time, entrenching the success of the demand-driven approach. However, this transition could be punctuated by occasional sharp liquidity crunches that the Fed might feel forced to avoid by building a larger standing supply of reserve balances. This could in turn detract from expectations that liquidity would

[31]Data: Federal Reserve Bank of New York, Repo Operations.

[32]I participated in this "Pop-up Symposium on Money Market Developments" held on November 17, 2025, co-hosted by Morgan Stanley and the Bank Policy Institute.

[33]This additional operation was temporary at the end of 2024 Statement Regarding the Standing Repo Facility from December 30, 2024, Federal Reserve Bank of New York. It was then made permanent on June 25, 2025 (Federal Reserve Bank of New York).

[34]Each eligible firm can obtain up to $120 billion in each of two operations per day, provided the firm submits the maximum proposition of $40 billion in each eligible type of operation, based on whether the collateral for the operation is (i) U.S. Treasuries only, (ii) Agency debt or U.S. Treasuries, and (iii) Agency mortgage backed securities, Agency debt, and U.S. Treasuries (Federal Reseve Bank of New York.

[35]Data: Bank of England Weekly Report.

be supplied by the Fed only on a demand-driven basis. At this point, the Fed is taking an approach somewhere in between demand-driven and supply-driven, by encouraging the use of SRPs while also using reserve management purchases (RMPs) to keep rate volatility in check.

5 Temporary open market operations

Under my central hypothesis that the minimum quantity of reserve balances is that necessary to meet the needs of the payment system, shortfalls to that minimum contribute to less effective monetary policy transmission and to the potential for liquidity crunches. For example, when increases in the TGA generate unintended shortfalls in the supply of reserves, the Fed could sterilize those supply shocks with offsetting temporary open market operations (TOMOs), as suggested by Nelson (2024) and Vissing-Jorgensen (2025).

The same principle can be applied when the supply shocks are generated by increases in balances held in the Fed's Foreign and International Monetary Authorities (FIMA) reverse repo pool[36] or by the end-of-period window dressing conducted by foreign banks that are shedding reserve balances to reduce their capital requirements. The TGA and the FIMA reverse repurchase pool are the two largest "Factors Affecting Reserve Balances of Depository Institutions, other than paper currency. As of the end of February, 2026, the TGA balance is $887 billion and the FIMA reverse repurchase pool is $321 billion.

Unintended end-of-period shocks to the supply of reserve balances tend to disrupt monetary policy transmission, as discussed in the Introduction. After controlling for other determinants of SOFR-IORB spreads, Copeland, Duffie, and Yang (2025a) estimate a quarter-end fixed effect on SOFR-IORB spreads of about 10 basis points for the sample period 2015-2023, and about 30 basis points when predicting SOFR-IORB at the 99th percentile—that is, in potentially stressed market conditions, when the loss of reserve balances matters most.[37] Correa, Du, and Liao (2025) and Kloks, Mattille, and Ranaldo (2025) also find significant end-of-quarter impacts on repo-rate spreads and, as a consequence, on the cross-currency basis.

Supplying significantly *more* reserve balances than necessary to run the payment system can also interfere with monetary policy transmission. This fact motivated the Fed to introduce the Overnight Reverse Repurchase Facility, which absorbs excess reserve balances whenever these excesses push money market rates below the Fed's target. When the Fed prefers to have a smaller balance sheet, it could instead use TOMOs to soak up these extra reserves. This approach obviously doesn't apply when the Fed is expanding the supply of reserves to support quantitative easing or to mitigate dysfunction in Treasury markets (Duffie

[36]This is not to be confused with the "FIMA repo facility," whose quantities do not directly impact the supply of reserve balances. As for the FIMA reverse repurchase pool: "The New York Fed is authorized and directed by the Federal Open Market Committee (FOMC) to provide the FIMA reverse repo pool. This facility has existed since the mid-1970s. The FIMA reverse repo pool impacts the Federal Reserve's balance sheet. As account holders' demand for FIMA reverse repo pool investments increases or declines, the composition of the Federal Reserve's liabilities shifts, resulting in a corresponding reduction or increase in the U.S. banking system's reserves" (Federal Reserve Bank of New York). Data for the amounts held in the FIMA reverse repuchase pool are available from FRED.

[37]See Tables 1 and 2 of Copeland, Duffie, and Yang (2025a).

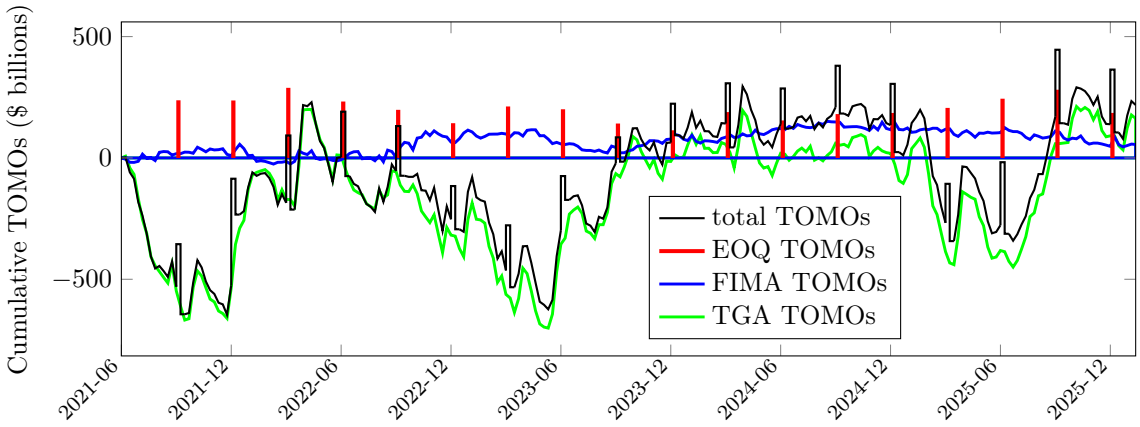
and Keane, 2023).

Figure 14a shows the cumulative quantity of TOMOs that would have been necessary, over the period 2021 to 2026, to counteract shocks to the supply of reserve balances caused by quarter-end window dressing by foreign banks and by fluctuations in balances held in the TGA and the FIMA reverse repurchase pool. The TOMOs that offset the predicted quarter-end shocks are reversed immediately after the end of each quarter, as depicted in Figure 14a.

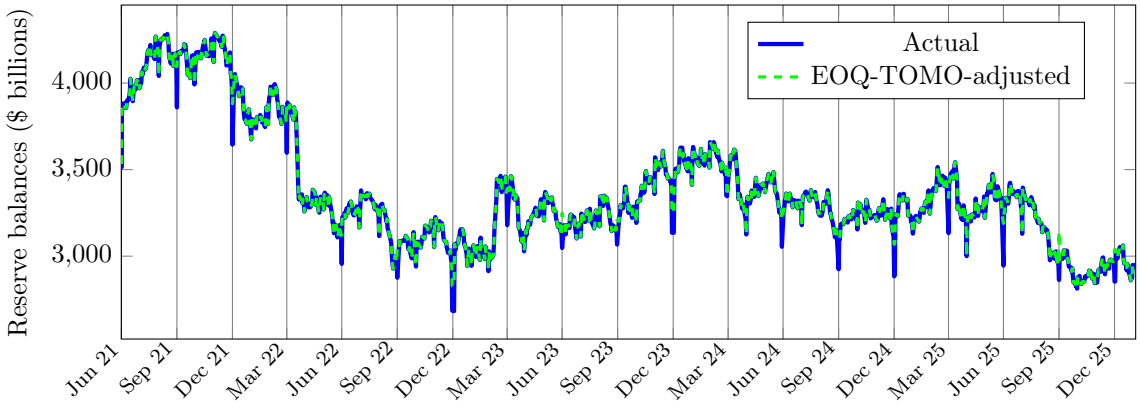
The quantities of TOMOs needed to offset period-end supply shocks are not fully predictable. However, if the approach is useful in principle then it would also be useful even if the prediction of quarter-end reserve balance supply shocks are imperfect. For the illustrative purpose of Figure 14a I chose a simplistic model of these TOMOs, which in practice would probably be sized with a more sophisticated model. I restrict attention to period ends that are quarter ends, whereas there are notable end-of-month effects for October and April.[38] I assume for illustrative purposes that the Fed's prediction of an end-of-quarter shock to reserve balances is the average of the shocks at the previous two quarter ends, before adjustment for TOMOs. I applied an exception to end-of-year TOMOs, based on the empirical evidence that year-end window-dressing shocks to reserve balances tend to last for two days rather than one, and tend to be larger than those at other quarter ends. Correspondingly, the illustrative year-end TOMOs are reversed after two days rather than one and are sized simply by scaling the average of two previous quarter-end supply shocks by the ratio of sample average year-end supply shocks to the average of the supply shocks for other quarter ends.

Using TOMOs to offset some of the unintended shocks to the supply of reserve balances can complement both the supply-driven and demand-driven approaches to achieving ample reserves. An argument against the use of TOMOs for this purpose is that this policy does not support a key principle of the demand-driven approach, which is that banks and their affiliated dealers will demand liquidity from the Fed only if they are given sufficiently strong rate incentives to do so. Over time, these rate incentives may generate enough use of the Fed's liquidity facilities that the associated stigma will be reduced, thus eventually reducing rate volatility and the risk of liquidity crunches. The TOMO approach that I have described would reduce rate volatility and thus reduce these demand-driven incentives. However, TOMOs that sterilize unintended bumps in the supply of reserves could be combined with a reduction in the average path of reserve balances over time that is sufficient to restore incentives supporting a demand-driven approach.

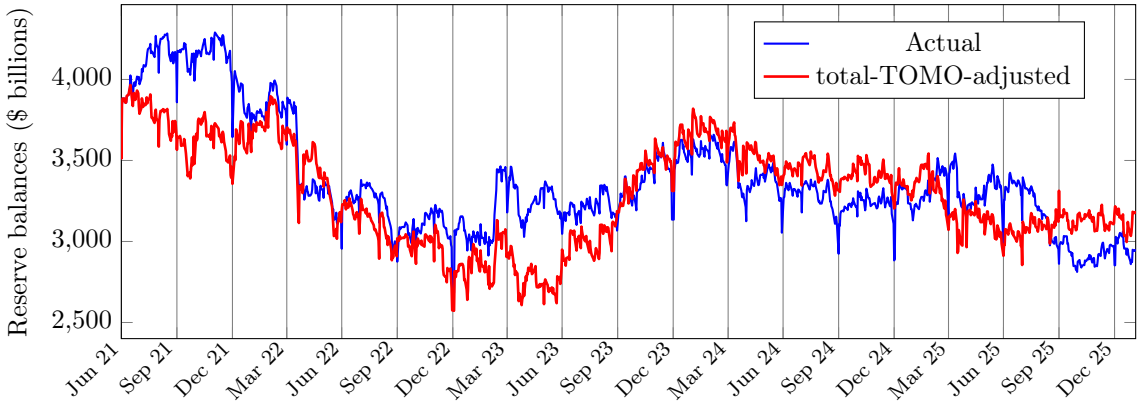
[38] At the end of October 2025, for example, Canadian banks with foreign branch office accounts at the Fed are monitored for capital sufficiency. Figure 1 shows the stress in the repo market at this month end.



(a) Cumulative TOMOs offsetting the impacts on actual reserve balances of end-of-quarter drops in reserve balances caused by foreign-bank window dressing (EOQ TOMOs), by changes in FIMA reverse repurchase pool (FIMA TOMOs), and by changes in TGA balances (TGA TOMOs). The time series are weekly. Non-year-end TOMOs are the average of the previous two EOQ drops in actual reserve balances, and are reversed on the next business day. Year-end TOMOs are reversed in two business days and are scaled by the ratio of average year-end drops in actual reserve balances to average non-year-end EOQ drops in actual balances.



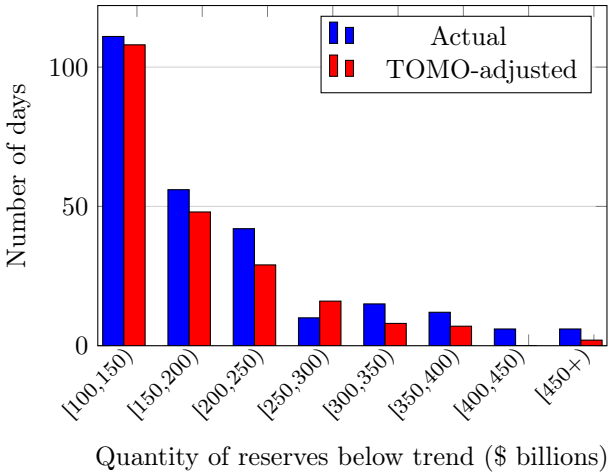
(b) Reserve balances before and after adjustment for the EOQ TOMOs shown in Figure 14a.



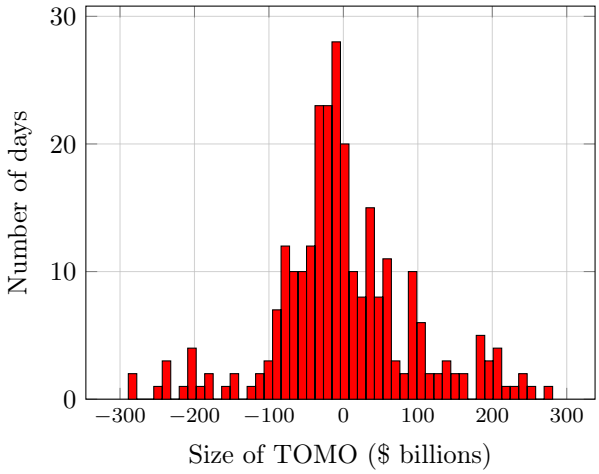
(c) Reserve balances before and after adjustment for the total TOMOs shown in Figure 14a

Figure 14: Hypothetical TOMOs and their impacts on reserve balances. Data: Actual reconstructed daily reserve balances: Lou Crandall, Wrightson Associates. TGA and FIMA repurchase pool balances: FRED.

Figure15a shows the implications for reserve balance shortfalls of the hypothetical TOMOs illustrated in Figure 14. Even though these TOMOs reduce many of the largest shortfalls relative to a 40-day moving average of past levels of reserves, there still remain substantial shortfalls that could generate rate volatility whenever reserves are close to minimally ample. The largest shortfalls to trend are for April 19 and 20, 2022, almost certainly caused by tax payments that convert reserves to TGA balances, and for December 31, 2021, when the modeled TOMO was small relative to the actual year-end drop in reserve balances. Figure15b illustrates the cross-sectional distribution of the sizes of the hypothetical TOMOs.



(a) Histogram of days on which the quantity of reserve balances, actual and TOMO-adjusted, was below trend by amounts in the indicated intervals.



(b) Histogram of the sizes of daily temporary open market operations, excluding days with no TOMOs.

Figure 15: Histograms of (a) days on which the quantity of reserve balances, actual and TOMO-adjusted, was below trend by at least $100 billion and (b) sizes of temporary open market operations. Trend reserve balances are defined as the average of the trailing 40 business days other than quarter ends. A positive TOMO increases the supply of reserve balances by the indicated amount. The sample covers 1329 business days. The TOMO sizes in panel (b) are first differences of the cumulative TOMOs shown in Figure 14c.

TOMOs that would sterilize the impacts of the TGA and the FIMA reverse repurchase pool could be executed with frequent (say, daily or weekly) Treasury repos and reverse repos. Some base amount of these TOMOs could be substituted with outright purchases and sales of Treasury bills (assuming that the Fed has a sufficient inventory of T-bills when taking reserves out of the market.)

Period-end TOMOs, however, are lumpier and less predictable. This may suggest the design of period-end TOMOs in the form of SRPs with special terms designed to generate large injections of reserve balances across the period end. For example, SRPs are normally priced at 10 basis points above IORB, but for period-end TOMOs, the Fed could perhaps accept bids as low as, say, 5 or 10 basis points *below* IORB. This could stimulate substantial demand and may also signficantly reduce the stigma that GSIBs and their observers normally associate with SRPs. After all, accepting funding at rates below IORB is an unusual profit opportunity that observers should not interpret as a signal of financial difficulty. Likewise, the arrival of the end of a quarter is not a surprise. Further, this sort of end-of-period SRP design does not seem to create moral hazard. Special end-of-period SRP funding improves

the transmission of monetary by offseting a known calendar effect in the supply of reserves caused by foreign-bank window dressing. This is not a "liquidity bailout."

Changes in the supply of reserves caused by fluctuations in the demand for paper currency are usually small. In March and April of 2020, however, the COVID pandemic caused an unusually large expansion in the supply of paper currency, by about $94 billion FRED. When reserves are not abundant, a TOMO could also be used to offset this sort of supply shock. In March and April of 2020 the Fed was already adding rapidly to the supply of reserves through purchases of Treasury securities designed to address dysfunction in the Treasury market.

The hybrid approach of using TOMOs to counteract predicted unintended changes in the demand or supply of reserve balances is a useful compromise if the Fed wants to reduce its balance sheet and is concerned that a pure demand-driven approach would result in excessive rate volatility and entail too much risk of liquidity crunches. This approach would be consistent with demand-drive reserves ampleness and also with a smaller Fed balance sheet if accompanied by a path of reserve balances that is lower on average over time. With a lower and smoother path of reserve balances, it should be possible to implement effective monetary policy transmission by relying primarily on adjustments in IORB.

6 Liquidity regulations

Post-GFC liquidity regulations have forced large banks to demonstrate that they can meet their liquidity needs from their own sources. This has consequences for the effectiveness of the payment system, given the reduced incentives for large banks to rely on central bank sources of liquidity, including daylight overdrafts, the Discount Window, and the Standing Repo Facility. In the analogous U.K. setting, Davey and Gray (2014) wrote that this "regulatory change has also had the unavoidable effect of incentivizing banks to economize on their intraday liquidity requirements for payment systems (so that they can reduce the size of their liquid asset buffer requirement). As discussed above, a CHAPS settlement bank uses its own funds when, at any point during the day, it has sent more payments than it has received. To reduce its liquidity requirement (that is, its need for own funds), a CHAPS settlement bank can therefore simply wait to receive payments from others before it sends payments. This behavior is referred to as being 'receipt-reactive."'

The Federal Reserve Board's Regulation YY, Enhanced Prudential Standards, includes rules that require globally systemically important bank holding companies (GSIBs) to be able to cover their intraday liquidity needs.[39] A 2026 poll of large banks conducted by Nelson (2026a) showed that the two most important factors driving the demand for reserves are

[39]The language for this rule in the Code of Federal Regulations includes: "If the bank holding company is a global systemically important BHC, Category II bank holding company, or a Category III bank holding company, these procedures must address how the management of the bank holding company will: (i) Monitor and measure expected daily gross liquidity inflows and outflows; (ii) Manage and transfer collateral to obtain intraday credit; (iii) Identify and prioritize time-specific obligations so that the bank holding company can meet these obligations as expected and settle less critical obligations as soon as possible; (iv) Manage the issuance of credit to customers where necessary; and (v) Consider the amounts of collateral and liquidity needed to meet payment systems obligations when assessing the bank holding company's overall liquidity needs."

(a) liquidity risk management (that is, ensuring that the bank has sufficient reserves to run its business) and (b) meeting the liquidity regulation known as Reg YY. The Fed's Large Institution Supervision Coordinating Committee (LISCC) supervises the intraday liquidity risk of large banks. In its May, 2019 Report on Supervisory Developments, the Federal Reserve Board stated: "In 2019, LISCC liquidity supervision is focusing on the adequacy of a firm's cash-flow forecasting capabilities, practices for establishing liquidity risk limits, and measurement of intraday liquidity risk." Ihrig (2019) describes the associated Comprehensive Liquidity Analysis and Review (CLAR). These CLAR reviews include an advantage to a GSIB of meeting its liquidity requirements by holding reserve balances, rather than in the form of other high quality liquid assets such as Treasury securities (Andolfatto and Ihrig, 2019; Bush and others, 2019).[40] This preference for reserve balances arises mainly from the fact that reserve balances are an instant source of liquidity, whereas converting securities to cash takes time. Bush and others (2019) note that "rapidly turning very large quantities of assets—even Treasury securities—into cash could be challenging. One problem is operational, as it might be difficult to find counterparties willing to purchase or repo unusually large quantities of assets on the same day an outflow occurs. Another issue is that potential counterparties may perceive an attempt to monetize a large quantity of assets as a signal of stress and, in response, hold on to their cash in case they need it later. ... An alternative to monetizing securities is for banks to hold reserves in their central bank accounts. Reserves don't need to be turned into cash, since they are cash, and are readily available to meet sudden outflows. Moreover, they are not constrained by the closing time of securities settlement systems." The preference of supervisors or GSIBs themselves for meeting liquidity regulations with reserve balances significantly increases the total demand for reserve balances. The analysis of Bush and others (2019) suggests that this regulatory source of demand for reserves may be in the hundreds of billions of dollars.

Separately, responding to the Dodd-Frank Act, the Fed and FDIC implemented failure planning requirements for Resolution Liquidity Adequacy and Positioning (RLAP), which include an intraday "resolution" liquidity requirement. The associated FDIC and Federal Reserve Board guidance states that during a failure-resolution situation banks must "ensure that liquidity is readily available to meet any deficits." "Additionally, the RLAP methodology should take into account (A) the daily contractual mismatches between inflows and outflows; (B) the daily flows from movement of cash and collateral for all inter-affiliate transactions; and (C) the daily stressed liquidity flows and trapped liquidity as a result of actions taken by clients, counterparties, key FMUs,[41] and foreign supervisors, among others."

Given these post-crisis liquidity regulations and supervision, GSIBs are highly averse to the risk that their intraday reserve balances could approach zero, the "red line" described

[40] Andolfatto and Ihrig (2019) write: "Since the financial crisis, banks are now using reserves to help meet liquidity regulations, such as the liquidity coverage ratio (LCR) and resolution planning. While U.S. Treasuries are given equal weight with reserves in the calculation of high-quality liquid assets (HQLA) for the LCR, they are evidently not considered equivalent for resolution purposes. Internal liquidity stress tests apparently assume a significant discount on Treasury securities liquidated in large volumes during times of stress, so that Treasuries are not treated as cash-equivalent. We have heard that banks occasionally feel under supervisory pressure to satisfy their HQLA requirements with reserves rather than Treasuries.

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

by JP Morgan CEO Jamie Dimon following the liquidity crunch of September 2019.[42] To a question about why JP Morgan did not invest more of its reserve balances in Treasury repos, at rates far above IORB, Dimon replied:

"As I said, we have $120 billion in our checking account at the Fed, and it goes down to $60 billion and then back to $120 billion during the average day. But we believe the requirement under CLAR and resolution and recovery is that we need enough in that account, so if there's extreme stress during the course of the day, it doesn't go below zero. If you go back to before the crisis, you'd go below zero all the time during the day. So the question is, how hard is that as a red line? Was the intent of regulators between CLAR and resolution to lock up that much of reserves in the account with Fed? And that'll be up to regulators to decide. But right now, we have to meet those rules and we don't want to violate anything we've told them we're going to do."

Banks also have access to Fedwire daylight overdrafts as an intraday source of additional reserve balances. Under the Fed's Payment System Risk Policy, "Reserve Banks provide intraday balances by way of supplying temporary, intraday credit to healthy depository institutions. The PSR policy recognizes that the Federal Reserve has an important role in providing intraday balances and credit to foster the smooth functioning of the overall payment system. Daylight overdrafts enable an institution to send payments more freely throughout the day than if it were limited strictly by its available intraday funds balance, increasing efficiency and reducing payment system risk."[43]

In practice, however, a daylight overdraft by a large systemically important bank could be interpreted by regulatory supervisors as a demonstration that the bank did not meet regulations requiring the bank to have self-sufficient liquidity resources. The very small daily peak level of total system daylight overdrafts in recent years, averaging under $5 billion per day in the last quarter of 2025,[44] suggests that GSIBs are highly averse to using this facility. Given the uncertain timing of incoming payments, GSIBs therefore have an incentive to delay their outgoing payments whenever there is a significant risk that incoming payments will be delayed. Faced with this potential risk, a GSIB has a relatively inelastic demand for reserves balances, up to the level at which beginning-of-day balances and incoming payments are likely to be sufficient to process payments without reliance on liquidity sourced from the Fed. Under these conditions, when a dealer bank has relatively little expected slack in its reserves balances it may fail to provide funding to others except at interest rates substantially higher than IORB (Correa, Du, and Liao, 2025; Copeland, Duffie, and Yang, 2025a).

The tiny recent amounts of daylight overdrafts by top-ten banks documented by Badev and others (2021) contrasts with much larger pre-GFC overdrafts by top-ten banks on the

[42]Dimon's comments during J.P. Morgan's third-quarter 2019 earnings call were covered by Bloomberg. Glenn Schorr, analyst at Evercore, questioned Dimon as follows. "Curious your take on everything that went on in the repo markets during the quarter, and I would love it if you could put it in the context of maybe the fourth quarter of last year. If I remember correctly, you stepped in the fourth quarter, saw higher rates, threw money at it, made some more money, and it calmed the markets down. I'm curious what's different this quarter that did not happen, and curious if you think we need changes in the structure of the market to function better on a go-forward basis."

[43]See Frequently Asked Questions: Temporary Actions to Support the Flow of Credit to Households and Businesses by Encouraging Use of Intraday Credit.

[44]See Data: Peak Daylight Overdrafts–Quarterly.

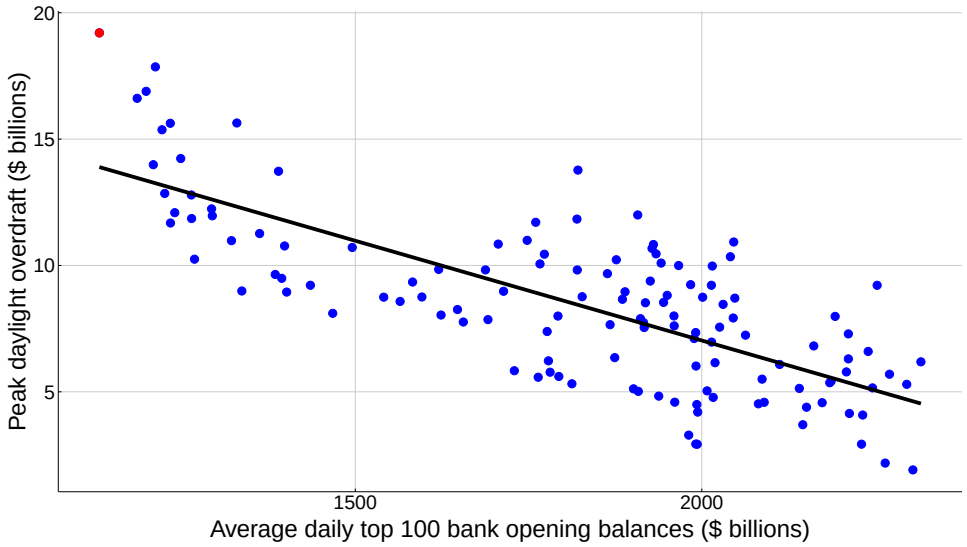


Figure 16: Peak intraday overdrafts and reserve balances Peak intraday overdrafts are calculated over two-week periods and published by the Federal Reserve. Average total reserve balances are computed over the same two-week periods. The R^2 of the linear relationship, plotted, is 0.57. The slope coefficient, -0.0079, is estimated with a standard error of 0.00061. The red dot corresponds to the observation for the 2-week maintenance period ending September 25, 2019. The data period is from January 1, 2015 to October 30, 2020. Figure source: Copeland, Duffie, and Yang (2025b) Data sources: Federal Reserve and Fedwire Funds Service.

order of $120 billion, for what was then a substantially smaller financial system (Badev and others, 2021). Some banks still use daylight overdraft to meet some of their payment needs when their reserve balances are insufficient. Indeed, as shown in Figure 16, peak system-wide overdrafts are highly related to the opening-of-day reserve balances of the 100 largest banks in the post-2015 sample period of Copeland, Duffie, and Yang (2025a), with an R^2 of 0.57 for this relationship. Figure 16 also shows that system-wide peak daylight overdrafts achieved their record high level in the two-week maintenance window containing the liquidity crunch of mid-September 25, 2019. This is also the two-week maintenance window of the sample with the lowest average daily opening balances. Figure 24 of Badev and others (2021) shows that top-ten banks did not avail themselves of significant daylight overdrafts during the mid-September 2019 liquidity crunch, despite record delays in their incoming payments (Figure 10) and record-high repo-rate spreads (Figure 12).

Former Fed Vice Chair Randall Quarles remarked that[45]

"I do think that one of the drivers of that disruption in 2019 was our, not so much liquidity regulation, but liquidity supervisory policy, which put a pretty heavy thumb on the side of the scale of preferring reserves over Treasury securities in satisfying your liquidity obligations. And with that heavy of a demand for reserves, it meant that people weren't going to jump into the Treasury repo market in the way that we expected that they would in order to kind of equalize that quickly, because they valued reserves much more heavily than they valued treasuries in satisfying what, from a regulatory purpose, are .... Those are viewed as the

[45]See Beckworth and Quarles (Beckworth and Quarles).

same thing. They will both equally well satisfy your regulatory obligation to maintain high-quality liquid assets."

According to the Federal Reserve Board's August 2019 Senior Financial Officer Survey, "satisfying internal liquidity stress metrics, meeting routine intraday payment flows, and meeting potential deposit outflows were important or very important determinants" of banks' holdings of excess reserves. A survey of the Bank Policy Institute showed that over three-quarters of the banks to which the Regulation YY liquidity buffer is applicable indicated that this is an "important" or "very important" consideration.

In 2024, the Fed "clarified" that GSIBs are permitted to count their borrowing capacity at the Discount Window and Standing Repo Facility as a source of liquidity for the purpose of meeting Reg YY liquidity requirements, to the extent of their highly liquid asset (HLA) collateral held at these facilities.[46] The Fed could allow banks subject to Reg YY to include among their liquidity resources *both the HLA and non-HLA* collateral they hold at the Fed's liquidity facilities. Nelson (2026b) explains how this would reduce the precautionary demand by GSIBs for reserve balances to meet liquidity regulations, but cautioned that "making such a substantial change will take time; it will require careful consideration and thoughtful design. For example, if banks are not willing to borrow from the discount window except under severe strain, it would be counterproductive to count capacity as a source of liquidity because the steps banks would take to avoid borrowing — selling assets at fire-sale prices and pulling back from lending to other banks — are precisely the way liquidity strains at one bank can propagate to the rest of the banking system."

In summary, it seems that post-GFC liquidity regulations and their supervision have had the unintended consequence of increasing the quantity of reserve balances necessary to run the payment system. While the adverse impacts of post-crisis financial regulations were predicted by regulators, they were not forecasted to be significant (Committee on the Global Financial System and Markets Committee, 2015).

Gorton, Laarits, and Muir (2020) predicted that analogous inefficiencies would be caused by the liquidity coverage ratio (LCR) rule, which impinges on a bank's high quality liquid assets, such as reserves, as a liquidity backstop. Their analysis draws from the National Banking Era, when national banks ignored apparent arbitrages that would have required issuing new money because of the distortionary effect of the requirement to back private money issuance one-for-one with Treasuries. Unlike National Banking setting, however, modern GSIB intra-day liquidity regulations are not like the "last taxi" metaphor of Goodhart (2008), by which regulation traps a portion of the liquid assets of a bank, preventing their use when needed.[47] With respect to the quantity of reserves necessary to run the payment system, the main problem is that liquidity regulations and their supervision have reduced the willingness of large banks to *increase* their available reserves when necessary by drawing on central bank facilities. This reduction in the short-term flexibility of the supply

[46] See Frequently Asked Questions about Regulation YY.

[47] Goodhart (2008) argues for the ineffectiveness of quantity-based liquidity regulation using the "last taxi in the station" metaphor. "The weary traveler who arrives at the railway station late at night, and, to his delight, sees a taxi there who could take him to his distant destination. He hails the taxi, but the taxi driver replies that he cannot take him, since local bylaws require that there must always be one taxi standing ready at the station."

of reserves causes an increase in the required standing supply of outstanding reserves. This is the primary distinction between the supply-driven and demand-driven approaches to ample reserves that I outlined in the Introduction.

The stigma associated with the Fed's liquidity facilities is longstanding and has only been deepened by post-GFC liquidity regulations and their supervision. Although the Fed is aware of these concerns, curing this stigma may be difficult. This is a chicken-or-egg situation in which each GSIB will perceive lower stigma only after other GSIBs are actively drawing liquidity from the Fed's facilities.

7 Tiering the remuneration of reserves

When there are frictions in interbank lending, banks with more reserves than necessary to conduct their payments may find it attractive to keep the excess balances as an investment, earning IORB, rather than lend them to a reserves-constrained bank. The frictional costs of borrowing and lending include capital requirements and FDIC insurance fees (Banegas and Tase, 2016). FDIC insurance fees currently range up to 42 basis points for the largest GSIBs, on top of the special additional fee assessment of up to 13 basis points that followed the failure of SVB and other banks in March, 2023.[48]

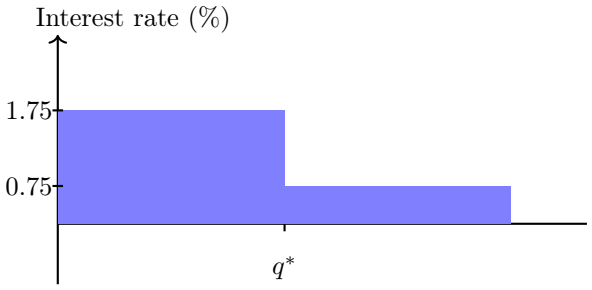
The Fed could minimize this sort of misallocation by providing reserves abundantly (so that most banks have low marginal liquidity benefits). Alternatively, rather than fixing a single remuneration rate, IORB, the Fed could reduce its remuneration once a bank hits some balance "quota." I will later discuss a "voluntary" quota approach that was developed by the Fed's Interest on Reserves Workgroup (2008). The difference between the two rate tiers should be near to or greater than the total of the frictional costs of borrowing and lending. With that, banks with excess reserves will lend to reserves-constrained banks, despite the frictional cost, given the relative marginal interest rates they are receiving from the Fed. I will outline a supporting theory.

For example, the central banks of New Zealand and Norway found it beneficial to tier their remuneration rates, as shown in Figure 17. I will discuss how this approach enlivened interbank lending in Norway and allowed the Reserve Bank of New Zealand to reduce the size of its balance sheet while keeping market rates stable.

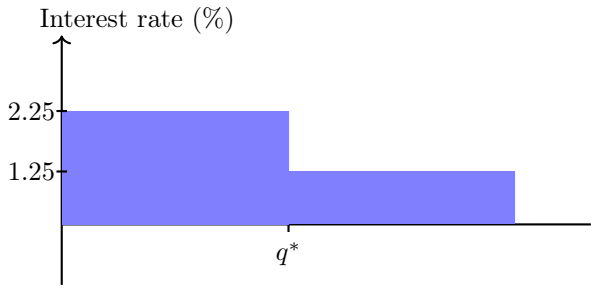
In April 2008, the Fed's internal Interest on Reserves Workgroup (2008)[49] provided a range of options to the FOMC for how to remunerate banks for reserve balances. One of these described the benefits of tiering the remuneration of reserve balances, under which "depository institutions could choose targets for their average balances over a relatively long maintenance period that might be set equal to the period between FOMC meetings. Each institution's average balance over the maintenance period would earn explicit interest at the target federal funds rate up to the upper end of a narrow clearing band – perhaps plus

[48]Source: Assessment Methodology and Rates, FDIC.

[49]This memo was confidential until 2015. The working group members were Jim Clouse (Co-Chair), Seth Carpenter, John Driscoll, Sherry Edwards, David Mills and Travis Nesmith from the Board; Spence Hilton (Co-Chair), Leo Bartolini, Chris Burke, Todd Keister, Antoine Martin and Jamie McAndrews (FRBNY); Ron Feldman (FRB Minneapolis); Steve Meyer (FRB Philadelphia); Huberto Ennis and John Weinberg (FRB Richmond).



(a) The RNBZ remunerated reserves below an assigned tier at the policy rate (e.g., 1.75% in December 2017) and above the tier at a lower level, typically the policy rate less 1%. See this paper for more information.



(b) The Norwegian Central Bank remunerates reserves below the quota at the policy rate (for example, 2.25% in October 2022) and above the quota at the reserve rate (for example, 1.25% in October 2022).

Figure 17: Tiered remuneration schemes used in New Zealand and Norway.

or minus 1 percent – around its target. Any amount by which an institution's average balance exceeded the upper end of its target band would be remunerated at a lower rate. The spread between the target federal funds rate and the rate paid on above-band balances would equal the spread between the primary credit rate and the target funds rate. If an institution's average balance fell short of the bottom of its target band, the deficiency would be penalized at a rate that would make the cost of falling short equal to or greater than the cost of borrowing at the primary credit rate." The Working Group explained that "Large banks likely would choose a target balance close to their estimates of their average need for balances to make payments without incurring overnight overdrafts."

The tiered remuneration approach considered by the Fed's Working Group has been theoretically analyzed by Whitesell (2006), Armenter (2016), Baughman and Carapella (2019), and Baughman and Carapella (2020). In their models, banks announce a target level of reserve balances. The central bank pays high interest on balances up to the target and low interest on balances exceeding the target. For incentive compatibility, the central bank assesses a fee on shortages relative to the target.

A key result of Baughman and Carapella (2019) is that tiering causes an increase in interbank lending activity. This is supported by the experience of Norway. Volumes in Norway's overnight interbank market rose dramatically after the Bank of Norway bank changed its policy for remuneration of reserve balances in October 2011 from a single tier to two tiers, as shown in Figure 18.

Why do some central banks tier their remuneration of reserves? Nield (2008) of the Reserve Bank of New Zealand wrote: "we considered that our role was to satisfy the demand for settlement account balances that were required for payment system purposes, rather than for investment demand. Therefore, in 2007 we introduced a pricing mechanism to discourage holdings of settlement balances beyond those needed for payment system purposes." In a Federal Reserve Board discussion paper Bowman, Gagnon, and Leahy (2010) wrote: "In response to the decline in demand for balances following the introduction in August 2007 of the tiered remuneration system, the RBNZ was able to reduce the overall quantity of settlement balances while keeping overnight market rates stable."

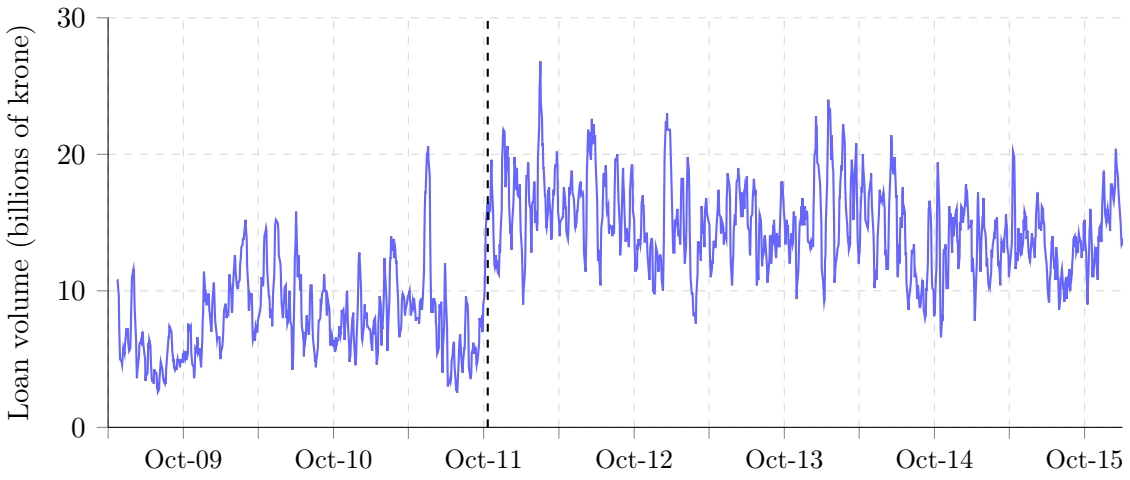


Figure 18: The total value of overnight interbank loans in billions of Norwegian kroner. The vertical dashed line corresponds to October 3, 2011, the date on which the Norwegian central bank changed its policy for remuneration of reserve balances from a single tier to two tiers. Balances above a "quota," determined separately for each bank. are compensated at a rate 100 basis points lower than balances above the quota. Figure source: Redrawn from Akram and Findreng (2021) with data provided by the authors.

7.1 Illustrative model of tiering benefits

The following simple model of tiered remuneration, summarized from Duffie, Sornwanee, and Spizzuocco (2026), illustrates how the quantity of reserve balances, total central bank interest expense, and total bank liquidity benefits can be traded off more effectively by using a tiered remuneration approach. The simplified version of the model described here assumes that banks are symmetric. The central bank first fixes its target i for the interbank equilibrium market borrowing rate. The central bank then implements this monetary policy by choosing a total quantity b of reserve balances and a marginal remuneration rate schedule $\{\rho(x) : x \geq 0\}$ that, in combination, imply an equilibrium market interbank borrowing rate equal to the target rate i. A marginal remuneration rate $\rho(q)$ at the level q of reserve balances implies that a bank with post-trade reserve balances q receives the total interest remuneration $\int_0^q \rho(x)\,dx$.

Before interbank borrowing and lending, the cross-sectional distribution of balances across the continuum of banks has a continuous cumulative distribution function F_b that is increasing in b, in the sense of first-order stochastic dominance. Each bank has a liquidity benefit $G(q, b)$ for its post-trade balance q that is increasing, differentiable, and concave in q, with $\lim_{q\to\infty} G_q(q, b) = 0$. That is, the marginal liquidity benefit to a bank of its reserve balances declines to zero as its balances go to infinity. The liquidity benefit of balances to a given bank, which depend on system-wide total balances b, include the ability to make timely payments and to avoid the actual or stigmatic costs of daylight overdrafts.

The frictional cost to a bank of borrowing an amount y in the interbank market is $c_b y$, for some proportional cost $c_b > 0$. Interbank lending of y involves a cost of $c_\ell y$ for some $c_\ell > 0$. Given a market interest rate i, a bank with initial balance x chooses to borrow some

amount $B(x)$, solving

$$\sup_{y \geq 0} \int_0^{x+y} \rho(z)\, dz - (i + c_b)y + G(x + y, b),$$

or to lend $L(x)$, solving

$$\sup_{y \geq 0} \int_0^{x-y} \rho(z)\, dz + (i - c_\ell)y + G(x - y, b).$$

The choices b and ρ of the central bank are consistent with the targeted market rate i provided that the market clears: $\int [B(x) - L(x)]\, dF_b(x) = 0$.

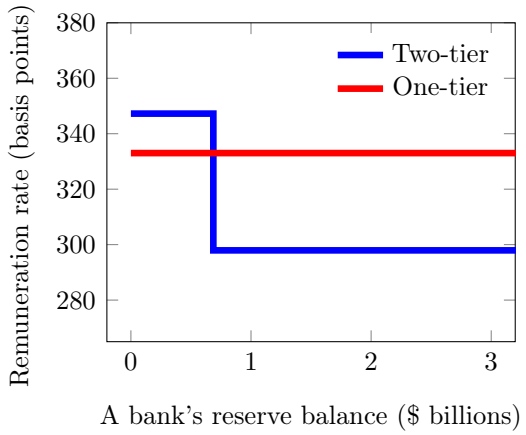
We assume that the central bank's interest expense is a key motivation, as suggested by Hughes and Younger (2025). The minutes of the Federal Open Market Committee (2018) reflect that interest expense is indeed a motive for a smaller Fed balance sheet. If the central bank restricts itself to a non-increasing remuneration schedule, Duffie, Sornwanee, and Spizzuocco (2026) show that total interest expense is minimized by the special case of a two-tiered remuneration schedule ρ, with

$$\rho(x) = \begin{cases} i + c_b - G_q(q_L, b), & x \leq q_L, \\ i - c_l - G_q(q_H, b), & x > q_H, \end{cases}$$

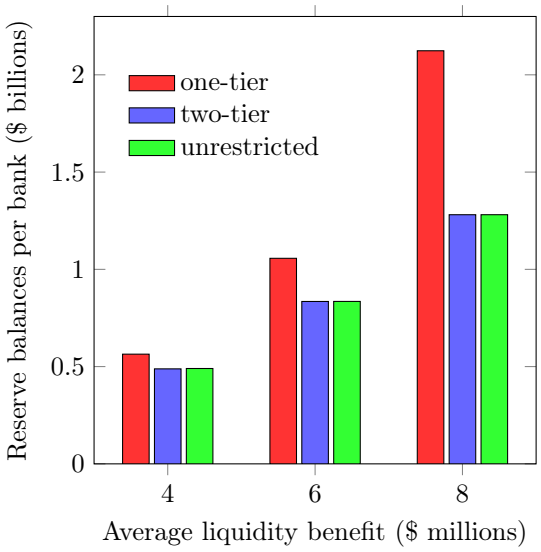
where, in equilibrium, q_L is the final balance of all borrowing banks and q_H is the final balance of all lending banks. (Market clearing implies that $F_b(q_L) = 1 - F_b(q_H)$.)

Figure 19 provides an illustrative example. As shown, the central bank can select various combinations of total reserve balances b and remuneration schedule ρ that are consistent with the targeted equilibrium interbank interest rate i. The average liquidity benefit[50] of banks depends on the choice of (b, ρ). The one-tier rate and the interest-expense-minimizing two-tier rate schedules for a given set of parameters are shown in Figure 19a. The unrestricted rate schedule that minimizes interest expense–which is not monotonic–can be found in Duffie, Sornwanee, and Spizzuocco (2026). The implications of all three types of remuneration schemes for total reserve balances, total central-bank interest expense, and total interbank lending volume are shown in Figures 19b, 19c, and 19d, respectively. As shown, there is typically a significant impact for central-bank outcomes of substituting a one-tier remuneration scheme with a two-tier scheme, but a much less important important effect caused by substituting a two-tier scheme with an unrestricted scheme.

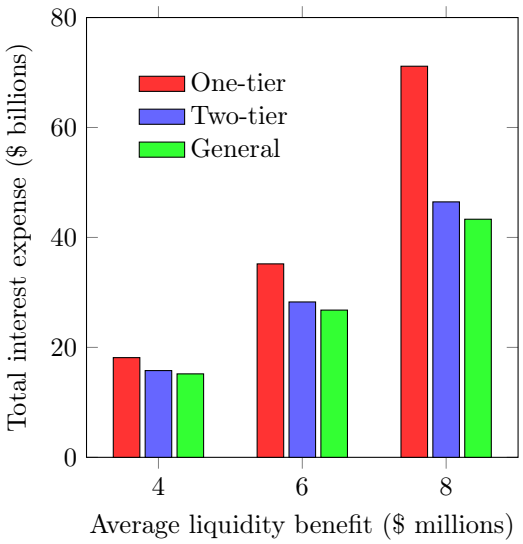
[50]The average post-trade liquidity benefit is $\int G(x + B(x) - L(x), b)\, dF_b(x)$. Duffie, Sornwanee, and Spizzuocco (2026) treat the frictional costs of trade as transfers rather than dead-weight social costs, given that in practice they are based mainly on capital requirements and deposit insurance fees.



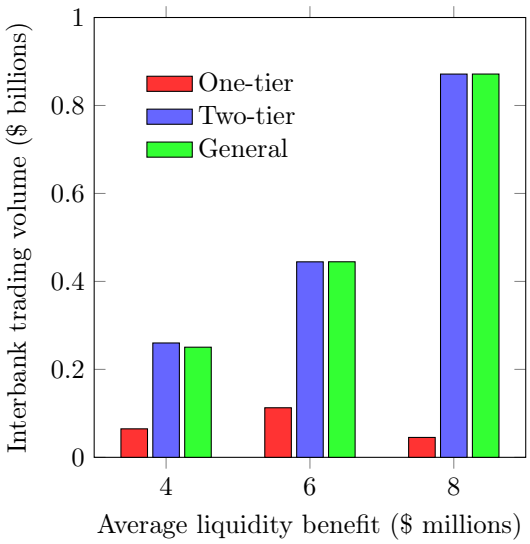
(a) The one-tier remuneration rate shown in red is 3.34%. The optimal two-tier remuneration schedule (blue) consistent with the targeted interest rate has a remuneration rate of 3.72% for balances up to a threshold of B_L = $1.35 billion, and a marginal remuneration rate of 3.09% for balances above B_L. This example assumes that system-wide total reserve balances are sufficient to imply an average equilibrium annualized liquidity benefit per bank of $6 million.



(b) The supply of reserve balances per bank for each of several cases of average bank liquidity benefit and for each form of remuneration schedule: one-tier (red), two-tier (blue), and unrestricted (green).



(c) The total interest expense of the central bank for each of several cases of average bank liquidity benefits and for each form of remuneration schedule: one-tier (red), two-tier (blue), and unrestricted (green).



(d) The total interbank borrowing transaction volume for each of several cases of average bank liquidity benefits and for each form of remuneration schedule: one-tier (red), two-tier (blue), and unrestricted (green).

Figure 19: Solution for an illustrative simple case with a target interbank equilibrium interest rate of 3.65%. The modeled liquidity benefit of a bank with a balance of q is $G(q,b) = k\left(1-e^{-\lambda q}\right)/\lambda$, where we take $k = 0.01$ and $\lambda = 0.8$. (Hence, we assume for this example, there is no dependence of a bank's liquidity benefit on system-wide total balances.) The cross-sectional distribution F_b of initial balances is log-normal with a mean that is by definition equal to system-wide total balances b and with a volatility (standard deviation of the underlying Gaussian) equal to $\sigma = 1$. The proportional frictional costs of borrowing and lending are $c_b = 0.004$ and $c_\ell = 0.003$, respectively.

In short, for this simple illustrative example, for a given average post-trade liquidity benefit of banks, when comparing with a constant remuneration scheme like IORB, a two-tiered remuneration scheme substantially increases interbank trading volume, reduces central bank interest expense, and reduces the quantity of reserve balances.

8 Liquidity savings mechanisms

Fedwire requires a large total amount of reserve balances to run smoothly. The timing of larger payments to the largest dealer banks, shown in Figure 20 suggests significant scope for reducing the required amount of reserve balances by augmenting Fedwire with a liquidity savings mechanism. If the Fed chooses to aim for a smaller balance sheet, an LSM would thus contribute to that objective. Moreover, with a demand-driven approach to reserve ampleness, the risk of a stressful liquidity event arising from an endogenous self-fulfilling bout of payment throttling is likely to be reduced by an effective LSM. However, the operational costs of establishing an LSM for Fedwire would be significant. Changing the software underlying the world's largest and most important payment system would be a major undertaking. The banks that would access such an LSM — perhaps only large banks — would also need to retool their own internal payment processing systems. It would probably take several years of significant systems work and testing to establish an effective LSM for Fedwire.

Kabadjova and others (2023) compare the relative efficiency of large value payment systems around the world. A key input for measuring payment system efficiency is the total "liquidity" demanded by the system, defined by $L = \sum_i L_i$, where L_i is the maximum intraday net outflow of balances of bank i (treated as zero if negative). Thus, L_i is the minimum sum of start-of-day reserve balances and daylight overdrafts that bank i would need to process its payments on that day. For Fedwire, the 2008-2018 sample average of L is \$629 billion. As a measure of the total amount of balances necessary to process payments, L is upward biased because banks do not all reach their intraday minimum balances at the same time. The measure of payment system efficiency proposed by Kabadjova and others (2023) is P/L, where P is the system-wide daily payment volume. The average payment efficiency for Fedwire over the study period is roughly 4, which is not unusual among the international panel of payment systems analyzed in this study. The Bank of England's large value payment system, CHAPS, has substantially higher average payment efficiency, exceeding 10 for most of the sample period, in part because of the effectiveness of the CHAPS LSM, which processes a bank's queued outgoing large payments by offsetting them with queued incoming payments, thus reducing the quantity of balances needed to make payments. Kabadjova and others (2023) find that payment efficiency varies with the degree of payment coordination among payment system participants, the total quantity of reserve balances (which typically has a negative effect on payment efficiency), the level of interest rates (which reflects the time value of delaying payments), incentives for early payment submission, and the use of liquidity saving mechanisms (LSMs).

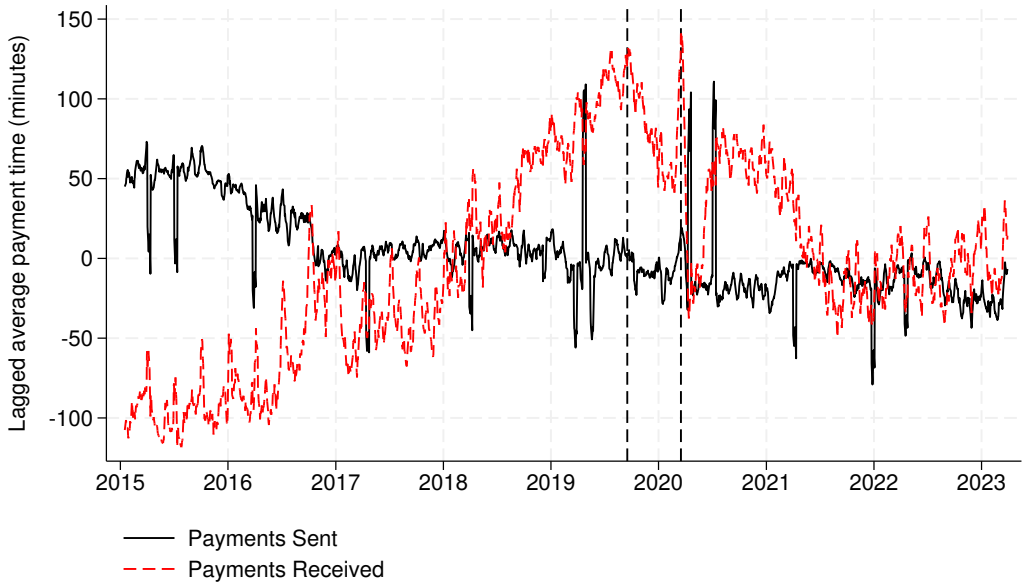


Figure 20: Timing of payments over $100 million to and from dealer banks. The times of day, relative to sample mean by which the dealer banks received half of their incoming payments and sent half of their incoming payments, relative to the sample mean, in minutes, for those payments of at least $100 million. (An increase in minutes means later in the day.) The observation for date t is the average over the past ten business days, inclusive of date t. The vertical lines mark September 17, 2019 and March 17, 2020. Data source: Fedwire Funds Service. Figure source: Copeland, Duffie, and Yang (2025b).

8.1 An LSM reduces the required total supply of reserve balances

An LSM allows a bank to send some of its payments by using little if any of its stock of reserve balances. Suppose, for an illustrative example, that Bank A begins with a reserve balance of 100 and will make a payment to Bank B of 60. Without an LSM, Bank A would make the payment from its reserve balances, leaving a balance of 40. With an LSM, Bank A would instead queue the payment for possible processing at the next intra-day LSM payment cycle. By the time of that LSM cycle, suppose that Bank B has queued an LSM payment to Bank C of 50 and Bank C has queued an LSM payment to Bank A of 55. Depending on its design, the LSM detects that it can process all three of these queued payments by reducing the balances of banks A and C by 5 each and by increasing the balance of Bank B by 10. Thus, the LSM allows Bank A to use only 5 of its balances rather than 60. By the end of the day, the LSM will have had no impact on the net amount of balances needed by Bank A. But the LSM can significantly reduce the liquidity needed by each bank, defined above as L_i, the maximum intraday draw-down in a bank's balances.

For the same illustrative payments, an alternative form of LSM could process partial payments of 50 to all three banks, leaving residual payments of 10, 0, and 5, respectively, to be made in the next LSM cycle.

Depending on the LSM design, little if any of the banks' initial reserve balances are needed to make these payments. This implies that the banks are able to start each day with a smaller amount of opening balances. The LSM would therefore allow the central bank to

achieve ample reserves with a smaller total supply of reserve balances.

How large a reduction in system-wide Fed reserve balances could be achieved by an LSM? Only a careful quantitative study could provide a useful estimate. Fedwire is the world's largest payment system and has a unique network topology. Payment volumes are dominated by the largest "money-center" banks. Eisenbach, Kovner, and Lee (2025) find that the five most active banks in Fedwire account for 50% of its payment volumes, with the 10 most active banks accounting for 60%. Fedwire includes a moderate number of large active regional banks and a very large number of smaller banks. However, Eisenbach, Kovner, and Lee (2025) show that, on a value-weighted basis, 80% of the potential links in Fedwire between payor and payee banks are completely inactive. To further complicate the calculation of potential LSM liquidity savings, Ashcraft and Duffie (2007) find that some money-center banks are net receivers in the early part of the day and then become net payers, thus requiring relatively low opening-of-day balances to process their payments. Other money-center banks, however, have a pattern of large outgoing payments followed by large net incoming payments, thus requiring relatively larger amount of reserve balances to process their payments. For JPMorgan, this intraday pattern is consistent with the remarks by Jamie Dimon quoted in Section 6.

The operational cost of establishing an LSM for Fedwire could be significant. Changing the software underlying the world's largest and most important payment system is no small matter. Moreover, the set of banks that would access such an LSM (perhaps only the largest banks) would need to retool their own intermal payment processing systems. It would probably some take years of significant systems work and testing to establish an effective LSM for Fedwire.

While there is little direct evidence at this point about how much an LSM would reduce Fedwire's liquidity demands, LSMs have been effective in other large value payment systems. Success depends on the extent of cycles (network payment loops) of potentially offsetting payments and on the incentives of banks to cede a significant fraction of their large payments to the LSM, rather than stay in unilateral control of their timing by making payments from their existing stock of reserve balances.

Norman (2010) notes that "the Bank of Korea's BOK-Wire+ payment system is estimated to have achieved liquidity savings in the order of 20% within a month of its launch, when compared to previous liquidity requirements in the BOK-Wire RTGS system," and $CHAT's efficiency increased from 8 before 2006 to 12 in 2007, an improvement of 50%. Norman (2010) also relates that liquidity savings of nearly 15% are estimated for banks in the Japanese large-value payment system, BOJ-Net, which is described later in this section.

Davey and Gray (2014) estimate that introducing an LSM into the Bank of England's large value payment system, CHAPS, led to a 20% reduction in required liquidity. Seaward (2016) discusses how the liquidity savings for CHAPS could, depending on the circumstances, be higher or lower than this estimate.

Chande (2023) estimate a payment efficiency of 10.2 for Canada's new large-value system, Lynx, compared to 7.6 for Canada's previous large value payment system, an improvement of over 30%, although these estimated efficiencies are somewhat noisy. They attribute this improvement to Lynx's LSM.[51] Canada's banks have the choice of which payments to queue

[51] Byck and Heijmans (2020) simulates the performance of various LSMs alone and in combination in the

in Lynx's LSM. A sign of the efficiency of this LSM is that over 99% of payments in Lynx are now settled in its LSM (Pasin and Wyllie, 2025). The degree to which Lynx queuing reduces the quantity of balances necessary to process payments has actually given Canada's banks an incentive to *reduce* the quantity of balances committed to the LSM in order to force the LSM to offset payments with each other rather than make payments from the balances available to the LSM (Garratt, Lu, and Tian, 2025).

An LSM is only effective if it provides sufficient incentives for banks to place a large fraction of their large outgoing payments into the LSM's queue, where they can potentially be offset by large incoming large payments. (The incentives of banks to place their payments into an LSM is modeled later in this section.) When this happens, the LSM reduces the quantity of balances necessary to make the payments in a timely manner.

8.2 Only large Fedwire payments matter

As I have explained and as shown in Figure 20, delays in large payments to the ten largest repo dealers are not largely offset by later payments of these large dealers to others. Copeland, Duffie, and Yang (2025a) show that this form of payment-timing stress predicts distortions in overnight Treasury repo rates (relative to IORB), especially in the tail of this interest rate spread. Smaller payments do not add stress. This is evidenced by Figure 21, which shows that outgoing payments of less than $100 million do not vary much when total system balances change, and moreover that outgoing smaller payments are largely offset by incoming smaller payments. Even on March 17, 2020, when the Figure shows a jump in payment delays after the announcement that COVID was a global pandemic, delays of smaller incoming payments seem to be mitigated by delays in outgoing payments. Even on March 17, 2020, when the Figure shows a jump in payment delays after the announcement that COVID was a global pandemic, delays of smaller incoming payments seem to be mitigated by delays in outgoing payments. LSMs are designed primarily to offset large payments with each other.

Davey and Gray (2014) relate that the LSM introduced by Bank of England's large value payment system reduced CHAPS banks' intraday liquidity requirements by around 20%, and "reduced incentives for banks to adopt adverse behaviors to economize on their intraday liquidity requirements."[52] These "liquidity requirements" are the intra-day reductions in reserve balances that are necessary to process the payments.

8.3 Case illustration: Japan's Liquidity Savings Mechanism

The Bank of Japan's large value payment system, the BOJ-NET Funds Transfer System (Bank of Japan, 2019) provides a helpful illustration of LSMs. Figure 22 is a flowchart of the handling of a payment between two financial institutions, Banks A and B. As shown, before anything else happens, the payment system software first checks whether the payment from Bank A to Bank B can be offset to some extent by an existing queued payment from Bank B to Bank A. This bilateral offset is illustrated in Figure 23a.

context of Canada's large value payment system, showing large potential reductions in liquidity requirements.

[52]For more background, see Seaward (2016).

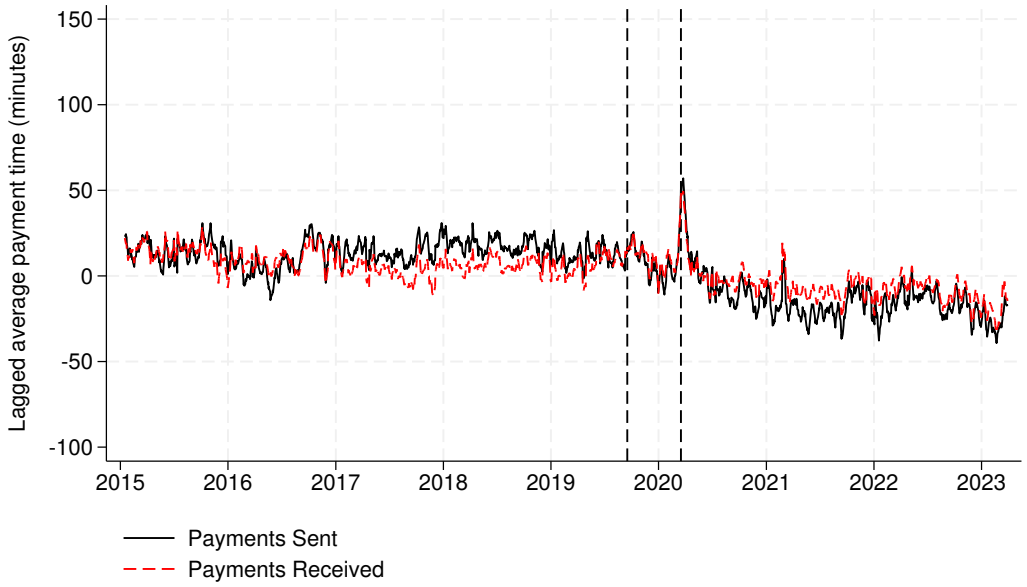


Figure 21: Timing of payments less than $100 million to and from dealer banks. The times of day, relative to sample mean by which the dealer banks received half of their incoming payments and sent half of their incoming payments, relative to the sample mean, in minutes, for those payments less than $100 million. (An increase in minutes means later in the day.) The observation for date t is the average over the past ten business days, inclusive of date t. The vertical lines mark September 17, 2019 and March 17, 2020. Data source: Fedwire Funds Service. Figure source: Copeland, Duffie, and Yang (2025b).

If a bilateral offset of the payment from A to B is not feasible, the system then tests whether the payment can be made with the existing balances of Bank A. If so, the payment is made immediately. If not, the payment is placed into a queue for later settlement.

At each of eight points in time during the day, a multilateral offsetting algorithm searches the current queue for feasible sequences of payments among multiple banks, like that shown in Figure 23b. At each of these eight cycles, this algorithm begins by determining whether all of the queued payments can be paid immediately. If not, the mechanism proceeds iteratively, round by round, successively removing infeasible payments at each round. There are many ways to do this. The method used by BOJ-NET removes the largest queued payment of each of participants that has insufficient balances to process its payments in the preceding round. This iterative pruning of payments is repeated, round by round, until all of the remaining payments can be made with the available balances. (There is a possibility that the last round leaves no feasible payments.)

8.4 LSM Theory

The timing of submission of payments into RTGS systems is subject to the preferences of banks. LSMs try to overcome coordination failures by which banks delay their payments because they worry that other banks will do the same. An LSM gives each bank an option to place its payments in a queue for potential early offsetting, as explained by Martin and McAndrews (2008), who provide an equilibrium theory. Duffie, Singh, and Wang (2026)

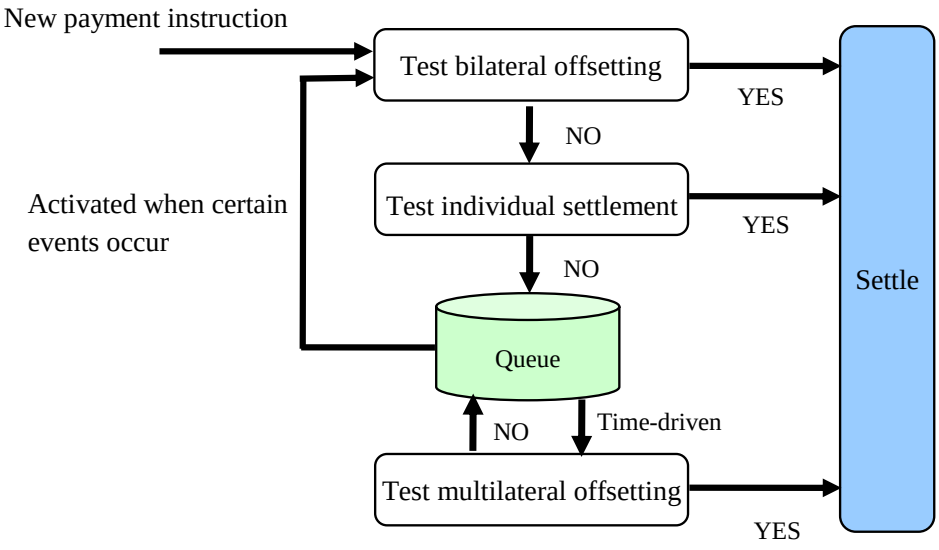
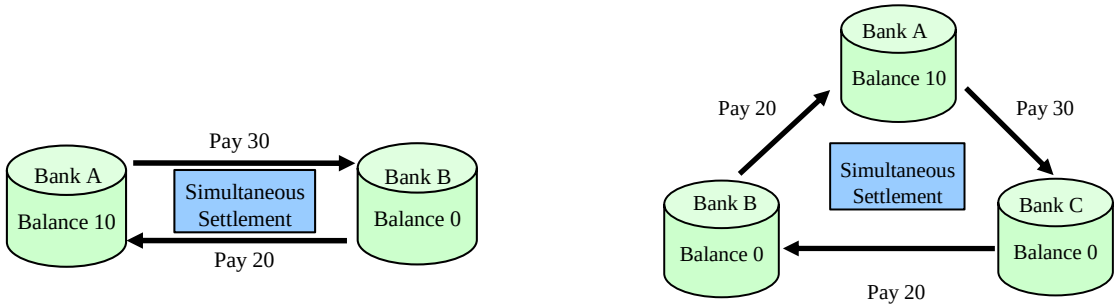


Figure 22: How BOJNET sequences its LSMs. Figure source: Bank of Japan (2019).



(a) Bilaterally offseting payments.

(b) Multilaterally offsetting payments.

Figure 23: Bilateral (a) and multilateral (b) payment offsets. Figure source: Bank of Japan (2019).

model an LSM that allows for full or partial submission of payments by a given bank to multiple other banks. I summarize a simplified version of their LSM here in order to illustrate the potential effectiveness of LSMs.

For $n \geq 2$ banks, the total required payment from bank i to bank j is $Q_{ij} \geq 0$. The beginning-of-day balance of bank i is X_i. The pairs $(X_i, \{Q_{ij}\})$ are assumed to be independent across banks. Payments can be made in one of two periods, early and late. Given early payments $p_{ij} \leq Q_{ij}$, bank i incurs an overdraft cost of $\psi(\sum_j p_{ij} - X_1 - \sum_j p_{ji})^+$, where $x^+ = \max(x, 0)$, and a late payment cost of $c(\sum_j Q_{ij} - p_{ij})$, for some positive coefficients ψ and c.

Yang (2020) analyses the equilibrium payment behavior of this setting without LSMs, and the implications for repo-rate distortions. We will compare his no-LSM equilibrium with that for the following LSM, analyzed by Duffie, Singh, and Wang (2026).

We let $N = \{1, \ldots, n\}$ and $G = \{(i, j) \in N^2 : i \neq j\}$. Bank i reports an upper bound

$\bar{p}_{ij} \leq Q_{ij}$ on the payment to bank j that bank i is willing to queue in the LSM. In practice, banks typically make payments on an all-or-none basis. All-or-none constraints, however, imply an extremely challenging combinatorial equilibrium analysis. Our analysis can be viewed as an approximation of the case in which Q_{ij} is the total of a number payments from i to j, only some of which are queued in the LSM.

Given the bank submissions, the LSM maximizes total payment volume by making the payments $\{p^*_{ij} : (i,j) \in G\}$ that solve the linear program

$$\begin{aligned} \max_{\{p_{ij}:(i,j)\in G\}} \quad & \sum_{i,j} p_{ij} \\ \text{s.t.} \quad & \sum_j p_{ij} - \sum_j p_{ji} \leq 0, \quad i \in N, \\ & 0 \leq p_{ij} \leq \bar{p}_{ij}, \quad (i,j) \in G. \end{aligned}$$

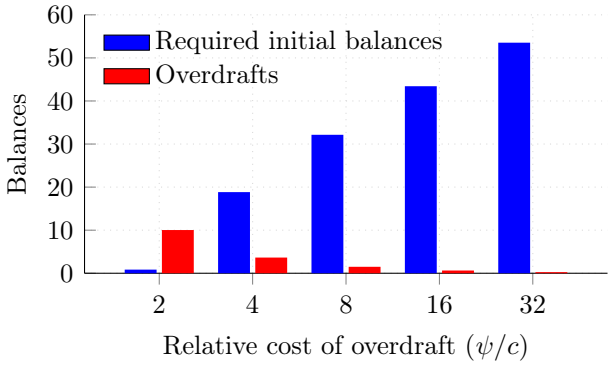
This LSM makes no payments out of balances, which are made instead by offsetting outgoing payments with incoming payments. Duffie, Singh, and Wang (2026) show that it is a weakly-dominant strategy for bank i to make the maximal submissions $\bar{p}_{ij} = Q_{ij}$. Duffie, Singh, and Wang (2026) show improved performance for LSMs that also incorporate payments of initial reserve balances.

For illustrative parameters, Figure 24b shows how the mean fraction of early payments depends on the mean opening balances of each bank, with and without an LSM. For this example, the proportional cost ψ of overdrafts is assumed to be twice the proportional cost c of late payments. Figure 24a shows the impact of increasing a bank's perceived cost of using daylight overdrafts on the liquidity resources required to achieve, in expectation, 50% early payments. As the cost ψ of overdrafts rises relative to the cost c of late payments, the required initial balances rises much more than the associated fall in expected quantity of overdrafts. (For the LSM to achieve 50% expected early payments, no initial balances are used to make payments for the illustrated case, so no overdrafts are required.) Figure 24a captures the idea that if banks are reluctant to overdraft because of liquidity regulations, the total quantity of reserve balances needed to run the payment system may rise significantly.

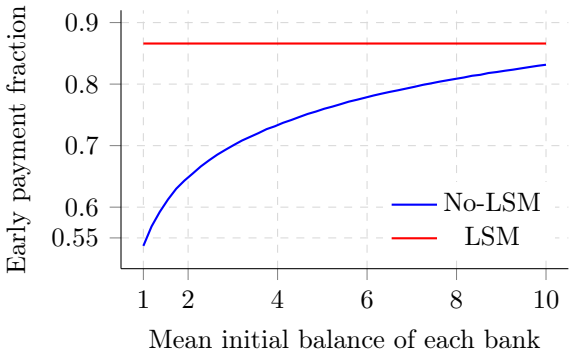
9 Policy summary and prioritization

Even though the quantity of reserve balances has grown by a factor of roughly 200 since 2007, the Fed is now increasing the supply of reserves to keep pace with the increasing demands of the payment system. Under post-GFC liquidity regulations, the Fed has found that it cannot fully rely on GSIBs with temporary needs for extra balances to use the Fed's on-demand liquidity, including daylight overdrafts, the Discount Window, and Standing Repo Operations.

If the Fed were to significantly reduce the size of its balance sheet, the supply of reserve balances would need to be reduced by about the same amount. If this were to be attempted with no changes in the Fed's operating framework and liquidity regulations, the disruption of monetary policy transmission could be severe. In extreme cases, a liquidity crunch could threaten financial stability. It has been difficult for the Fed to accurately predict the min-



(a) In order to achieve an expected fraction of 50% of early payments, the vertical axis shows required mean initial balances per bank (blue) and the mean chosen quantity of overdrafts per bank (red), depending on the relative cost of daylight overdrafts.



(b) How the mean fraction of early payments depends on the mean opening balances of each bank, with and without an LSM. The proportional cost ψ of overdrafts is twice the proportional cost c of late payments.

Figure 24: Panel (a) shows the impact of increasing a bank's perceived cost of using daylight overdrafts on the liquidity resources (opening balances and overdrafts) required to achieve a mean of 50% early payments. For the given parameters, the LSM achieves 50% expected early payments with no initial balances and without overdrafts. Panel (b) shows how the expected fraction of early payments, with and without an LSM, depends on the mean opening balance of each bank. The network consists of $n = 20$ symmetric banks. The payment Q_{ij} from bank i to bank i is $k_{ij}Z_{ij}$, where k_{ij} is binomial with equiprobable outcomes 0.1 and 1.9 and Z_{ij} is Poisson distributed with parameter $\gamma = 12$. The opening balance X_i of bank i is binomial with equiprobable outcomes $0.1\bar{X}$ and $1.9\bar{X}$, for some common mean $\bar{X}$. The variables k_{ij}, Z_{ij}, and X_i are independent. Figure source: Duffie, Singh, and Wang (2026).

imum supply of reserves necessary to avoid these outcomes. As a result, the Fed is now conservatively increasing total reserve balances. This might be the least costly policy for the Fed, although it could lead over time to ever greater reliance by banks on the outstanding stock of reserve balances to meet their payment and liquidity needs, a ratchet effect (Acharya and Rajan, 2022; Acharya, 2022).

If the Fed decides to reduce its balance sheet, it may be able to mitigate this situation with temporary open market operations that sterilize unintended shocks to the supply of reserve balances and by improving the mobility of reserves across financial institutions with: (1) changes in intraday liquidity regulations and their supervision; (2) adding a liquidity savings mechanism to Fedwire, perhaps combined with throughput rules or incentives; and (3) tiering the interest rate paid by the Fed to banks on their reserve balances, so that banks not to have reserve balances that they don't need for processing payments.

How would these enabling policies be prioritized?

It would be natural to start with a program of temporary open market operations that sterilize unintended changes to the supply of reserve balances. This approach would not require legislation or changes in regulation, and could be implemented at the direction of the FOMC. These TOMOs would not change the Fed's approach to rate control, which would continue to rely primarily on administered rates. Rate control would likely be more effective because the TOMOs would smooth the average path of reserve balances. To the extent that a smoother path of balances reduces repo rate volatility and thus lowers the incentives of dealers to use SRPs, the Fed could replace some of this incentive by pushing down the smoothed

path of reserve balances. Perhaps SRPs would be then be less stigmatized, although that outcome is hard to predict.

Of the remaining policy tools, changes in intra-day liquidity regulations and their supervision seem to be the next most easily implemented, although it's not clear how far to go in changing or relaxing these regulations and their supervision. The first place to look is whether to allow assets held as collateral at the Fed's liquidity facilities that are not highly liquid to count toward required liquidity for the purposes of meeting the internal liquidity stress tests (ILSTs) required by Regulation YY. It also makes sense to re-examine the preference in regulations and their supervision for meeting liquidity requirements with reserve balances, relative to Treasury securities.

Adding an LSM to Fedwire, or building a suitable new version of Fedwire, would be a large multi-year project. It seems natural to begin now with a cost-benefit analysis, which would probably entail preliminary design work. Eventually implementing an LSM would also require large banks to retool their internal payment systems.

Tiering the remuneration of reserve balances would also be a difficult multi-year project. There are two challenging issues here. The first is how to determine high-tier quotas for each type of bank. The voluntary reserves targeting approach of the Fed's Interest on Reserves Workgroup (2008) may allow these quotas to be chosen by banks themselves, given the anti-biasing incentives built into this approach, but that proposition has not been tested. The Fed could begin by doing some work on this design. The second major challenge to implementing tiered remuneration of reserves will be convincing other decision makers of the benefits. Tiering the remuneration of reserves may require legislation. The banking industry would likely get engaged, given the potentially significant implications for bank profitability.

References

Acharya, Viral, Rahul Chauhan, Raghuram Rajan, and Sascha Steffen (2022, August). Liquidity Dependence: Why Shrinking Central Bank Balance Sheets is an Uphill Task. Jackson Hole Symposium Paper, Federal Reserve Bank of Kansas City.

Acharya, Viral and Raghuram Rajan (2022, May). Liquidity, Liquidity Everywhere, Not a Drop to Use: Why Flooding Banks with Central Bank Reserves May Not Expand Liquidity. New York University, NBER Working Paper 29680.

Afonso, Gara, Darrell Duffie, Lorenzo Rigon, and Hyun Shin (2022). How Abundant Are Reserves? Evidence from the Wholesale Payment System. Federal reserve bank of new york staff report 1040, november, Federal Reserve Bank of New York.

Afonso, Gara, Domenico Giannone, Gabriele La Spada, and John C. Williams (2025). Scarce, Abundant, or Ample? A Time-Varying Model of the Reserve Demand Curve. Staff Report 1019, Federal Reserve Bank of New York, November.

Afonso, Gara, Anna Kovner, and Antoinette Schoar (2011). Stressed, not frozen: The federal funds market in the financial crisis. *Journal of Finance 66*, 1109–1139.

Akram, Farooq and Jon Findreng (2021). Norwegian interbank market's response to changes in liquidity policy. *Journal of Banking & Finance 125*, 106078.

Anbil, Sriya, Alyssa Anderson, and Benjamin Eyal (2026). Bankers' Banks and their Role in the Federal Funds Market. FEDS Notes, Federal Reserve Board, January.

Anderson, Alyssa and Dave Na (2024). The Recent Evolution of the Federal Funds Market and its Dynamics during Reductions of the Federal Reserve's Balance Sheet. FEDS Notes, Federal Reserve Board, July.

Andolfatto, David and Jane Ihrig (2019). The Fed and a Standing Repo Facility: A Follow-Up. Federal Reserve Bank of St. Louis, On the Economy Blog, April.

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

Armenter, Roc (2016). A Tractable Model of the Demand for Reserves Under Nonlinear Remuneration Schemes. Federal Reserve Bank of Philadelphia, Research Department, Working paper 16-35.

Ashcraft, Adam and Darrell Duffie (2007). Systemic illiquidity in the federal funds market. *American Economic Review, Papers and Proceedings 97*, 221–225.

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

Badev, Anton, Lauren Clark, Daniel Ebanks, Jeffrey Marquardt, and David Mills (2021). Fedwire Funds Service: Payments, Balances, and Available Liquidity. Finance and Economics Discussion Series, Working paper 2021-070, October.

Banegas, Ayelen and Manjola Tase (2016). Reserve Balances, the Federal Funds Market and Arbitrage in the New Regulatory Framework. Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board, Finance and Economics Discussion Series Discussion Paper 2016-079, September.

Bank of Canada (2022). An Overview of Lynx, Canada's High-Value Payment System. Bank of Canada, May.

Bank of England (2021). Liquidity Saving Mechanism User Guide. Bank of England, February.

Bank of Japan (2019). Information Disclosure Based on the Principles for Financial Market Infrastructures: The BOJ-NET Funds Transfer System. Bank of Japan, June.

Baughman, Garth and Francesca Carapella (2019). A Simple Model of Voluntary Reserve Targets with Tolerance Bands. Federal Reserve Board, Finance and Economics Discussion Paper 2019-060, July.

Baughman, Garth and Francesca Carapella (2020). Voluntary reserve targets. *Journal of Money, Credit and Banking 52*, 583–612.

Bech, Morten and Rod Garratt (2003a). The Intraday Liquidity Management Game. *Journal of Economic Theory 109*, 198–219.

Bech, Morten L. and Rod Garratt (2003b). The intraday liquidity management game. *Journal of Economic Theory 109*(2), 198–219.

Beckworth, David and Randall Quarles. Randal Quarles on Inflation, Balance Sheet Reduction, Financial Stability, and the Future of the Fed.

Berentsen, Aleksander, Alessandro Marchesiani, and Christopher Waller (2014). Floor systems for implementing monetary policy: Some unpleasant fiscal arithmetic. *Review of Economic Dynamics 17*(3), 523–542.

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

Bowman, David, Etienne Gagnon, and Mike Leahy (2010). Interest on Excess Reserves as a Monetary Policy Instrument: The Experience of Foreign Central Banks. Board of Governors of the Federal Reserve System International Finance Discussion Paper 996, March.

Buckle, Simon and Erin Campbell (2003). Settlement bank behaviour and throughput rules in an RTGS payment system with collateralised intraday credit. Bank of England, Working paper 209, December.

Bush, Ryan, Adam Kirk, Antoine Martin, Philip Weed, and Patricia Zobel (2019). Stressed Outflows and the Supply of Central Bank Reserves. Federal Reserve Bank of New York, Liberty Street Economics, February.

Byck, Shaun and Ronald Heijmans (2020). How Much Liquidity Would a Liquidity-Saving Mechanism Save if a Liquidity-Saving Mechanism Could Save Liquidity? A Simulation Approach for Canada's Large-Value Payment System. *Journal of Financial Market Infrastructures 9*(3).

Cenicola, Ashlyn, Melanie Friedrichs, Robert Mann, and Luke Olson (2025). Sizing the U.S. Repo Market. The OFR Blog, Office of Financial Research, December.

Chande, Nikil, Zhentong Lu, Hiru Rodrigo, and Phoebe Tian (2023). Three Things We Learned About the Lynx Payment System. Staff Analytical Note 2023-14, Bank of Canada.

Chien, YiLi and Ashley Stewart (2024). Bank Reserves since the Start of Quantitative Tightening. On the Economy Blog, Federal Reserve Bank of St. Louis, April.

Committee on the Global Financial System and Markets Committee (2015). Regulatory Change and Monetary Policy. CGFS Working Paper Number 54, Bank for International Settlements.

Copeland, Adam, Darrell Duffie, and Yilin (David) Yang (2025a). Reserves Were Not So Ample After All. *The Quarterly Journal of Economics 140*(1), 239–281.

Copeland, Adam, Darrell Duffie, and Yilin (David) Yang (2025b). Online Appendix for 'Reserves Were Not So Ample After All'. *The Quarterly Journal of Economics 140*(1), 239–281.

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

Copeland, Adam and Sarah Yu Wang (2024). The Dueling Intraday Demands on Reserves. *Liberty Street Economics*, October.

Correa, Ricardo, Wenxin Du, and Gordon Y. Liao (2025). Us banks and global liquidity. *Journal of Political Economy Macroeconomics 3*(4), 574–620.

Davey, Nick and Daniel Gray (2014). How has the Liquidity Saving Mechanism reduced banks' intraday liquidity costs in CHAPS? Bank of England, *Quarterly Bulletin*, Q2, pages 180-189.

Dolan, Simon and Matt Roberts-Sklar (2025). Learning to navigate bumps in the road: An update on the Bank of England's Sterling Monetary Framework Facilities and the recalibration of the Operational Standing Facility. Bank of England, December.

Duffie, Darrell (2022). *Fragmenting Markets: Post-Crisis Bank Regulations and Financial Market Liquidity*. Berlin: DeGruyter.

Duffie, Darrell and Frank Keane (2023). Market Function Asset Purchases. Federal Reserve Bank of New York Staff Report 1054, February.

Duffie, Darrell, Srisht Fateh Singh, and Chaojun Wang (2026). An efficient liquidity savings mechanism. Working paper, Stanford University.

Duffie, Darrell, Thanawat Sornwanee, and Francesco Spizzuocco (2026). Work in progress on tiered remuneration. Stanford University.

Duygan-Bump, Burcu and Jay Kahn (2026). The Central Bank Balance-Sheet Trilemma. FEDS Note, Federal Reserve Board, January.

Eisenbach, Thomas, Anna Kovner, and Michael Junho Lee (2025). When It Rains, It Pours: Cyber Risk and Financial Conditions. *Federal Reserve Bank of New York Economic Policy Review 31(1)*, 1–24.

European Central Bank and Bank of Japan (2017). Payment systems: liquidity saving mechanisms in a distributed ledger environment. Project Stella, European Central Bank and Bank of Japan, September.

Federal Open Market Committee (2018). Minutes. Board of Governors of the Federal Reserve Board, November.

Federal Open Market Committee (2019). Transcript of the Meeting of the Federal Open Market Committee held on April 30 to May 1, 2019. Federal Reserve Board.

Federal Reserve Bank of New York (2025). Statement Regarding Reserve Management Purchases Operations. Operating Policy, Federal Reserve Bank of New York, December.

Gagnon, Joseph and Brian Sack (2020). Recent Market Turmoil Shows That the Fed Needs a More Resilient Monetary Policy Framework. Peterson institute for international economics, *Realtime Economic Issues Watch*, june 18, Peterson Institute for International Economics,.

Garratt, Rodney, Zhentong Lu, and Phoebe Tian (2025). How banks create gridlock in payment systems to save liquidity: The case of canada. *Journal of Money, Credit and Banking*.

Gissler, Stefan, Samuel Hempel, Jay Kahn, Patrick McCabe, and Borghan Narajabad (2025). Monitoring Reserve Scarcity Through Nonbank Cash Lenders. FEDS Notes 2025-03-28-4, Board of Governors of the Federal Reserve System, March.

Goodhart, Charles (2008). Liquidity risk management. *Banque de France Financial Stability Review 11*(6), 39–44.

Gorton, Gary, Toomas Laarits, and Tyler Muir (2020, September). Mobile Collateral versus Immobile Collateral. NBER Working paper 22619.

Hamilton, James (1996). The Daily Market for Federal Funds. *Journal of Political Economy 104*, 26–56.

Hughes, Chris and Josh Younger (2025, September). Subsidy in the Shadows: Quantifying the Cost of Ample Reserves in U.S. Monetary Policy Implementation. Working paper.

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

Interest on Reserves Workgroup (2008). Interest on Reserves: A Preliminary Analysis of Basic Options. Federal Reserve Board, April 2008, Authorized for public release by the FOMC Secretariat on May 6, 2015.

Judson, Ruth and Elizabeth Klee (2009). Demand for Reserves on a Period Average Basis. Finance and Economics Discussion Series 2009-22, Federal Reserve Board, February.

Kabadjova, Biliana, Anton Badev, Saulo Bastos, Evangelos Benos, Freddy Cepeda-Lopéz, James Chapman, Martin Diehl, Ioana Duca-Radu, Rodney Garratt, Ronald Heijmans, Anneke Kosse, Antoine Martin, Thomas Nellen, Thomas Nilsson, Jan Paulick, Andrei Pustelnikov, Francisco Rivdeneyra, Mario Rubem do Coutto Bastos, and Sara Testi (2023). Intraday liquidity around the world. Bank for International Settlements, Working paper 1089, Monetary and Economic Department, April.

Kashyap, Anil K, Dimitrios P Tsomocos, and Alexandros P Vardoulakis (2024). Optimal bank regulation in the presence of credit and run risk. *Journal of Political Economy 132*(3), 772–823.

Kloks, Peteris, Edouard Mattille, and Angelo Ranaldo (2025). Hunting for Dollars. University of Basel, Swiss Finance Institute Research Paper 24-52, September.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2011). The effects of quantitative easing on interest rates: Channels and implications for policy. Brookings Papers on Economic Activity, 2011, pp. 215-87.

Leistra, Melissa (2026). Contrasting Ledgers: considerations for U.S. dollar interbank payment systems. Finance and Economics Discussion Series Working Paper 2026-011, Federal Reserve Board, February.

Logan, Lorie (2025). Ample liquidity for a safe and efficient banking system. Federal Reserve Bank of Dallas, October.

Logan, Lorie and Sam Schulhofer-Wohl (2025). Options for modernizing the FOMC's operating target interest rate. Federal Reserve Bank of Dallas, In Depth, September.

Martin, Antoine and James McAndrews (2008). Liquidity-Saving Mechanisms. *Journal of Monetary Economics 55*, 554–567.

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

McAndrews, James and Simon Potter (2002). Liquidity Effects of the Events of September 11, 2001. *Economic Policy Review 8*.

Miran, Stephen (2025). Regulatory Dominance of the Federal Reserve's Balance Sheet. Speech at At the Bank Policy Institute and Small Business & Entrepreneurship Council, Washington, D.C., Federal Reserve Board, November.

Nelson, Bill (2024). How the Federal Reserve Got So Huge, and Why and How It Can Shrink. Bank Policy Institute, February.

Nelson, Bill (2026a). What Determines Banks' Demand for Reserve Balances? Bank Policy Institute, January.

Nelson, Bill (2026b). Striking the right balance sheet. Testimony before the U.S. House Financial Services Committee Task Force on Monetary Policy, Treasury Market Resilience, and Economic Prosperity, Bank Policy Institute, January.

Nield, Ian (2008). Evolution of the Reserve Bank's liquidity facilities.

Norman, Ben (2010). Liquidity saving in real-time gross settlement systems – An overview. Bank of England, Financial Stability Paper Number 7, May.

Pasin, Virgilio and Anna Wyllie (2025). Net Send Limits in the Lynx Payment System: Usage and Implications. Bank of Canada, Staff Discussion Paper 2025-13, November.

Payments Risk Committee (2019). Best Practices for Payments, Clearing, and Settlement Activities. Federal Reserve Bank of New York, April.

Perli, Roberto (2025). Money Market Conditions and the Federal Reserve's Balance Sheet. 2025 U.S. Treasury Market Conference, Federal Reserve Bank of New York, November.

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

Remache, Julie (2026). Monetary Policy Implementation in an Ample Reserves Regime. Federal Reserve Bank of New York, Speech, February.

Ritchie, Greg and James Hirai (2026). Banks Borrow Record 100.9 Billion Pounds From BOE Short-Term Repo Tool. *Bloomberg*, January 26.

Saporta, Vicki (2025). Learning by doing. Bank of England, Speech given at Bank of Finland and SUERF Conference, Helsinki, June.

Schnabel, Isabel (2025). Towards a new Eurosystem balance sheet. Speech at the ECB Conference on Money Markets, European Central Bank, November.

Seaward, David (2016). Saving Liquidity in a Liquidity-Abundant World: Why don't banks use less liquidity when making high-value payments? Bank Underground, a blog for Bank of England staff, February.

The Clearing House (2025). CHIPS Liquidity Algorithm Unlocks Over $5 Billion in Annual Savings, Delivering Efficiency to the Global Financial Market. The Clearing House, press release, February.

Vissing-Jorgensen, Annette (2025). Fluctuations in the Treasury General Account and their effect on the Fed's balance sheet. FEDS Notes, Federal Reserve Board, August.

Waller, Christopher (2025). Demystifying the Federal Reserves's Balance Sheet. Federal Reserve Board, Speech at the Federal Reserve Bank of Dallas, July.

Warsh, Kevin (2025). Commanding Heights: Central Banks at a Crossroads. IMF Lecture, Group of 30, April.

Whitesell, William (2006). Interest rate corridors and reserves. *Journal of Monetary Economics 53*, 1177–1195.

World Bank Payment Systems Development Group (2020). Payment Systems Worldwide: A Snapshot. World Bank Global Payment Systems Survey, June.

Yang, Yilin (David) (2020). What Quantity of Reserves Is Sufficient? SSRN Working Paper.