How the Federal Reserve Got So Huge, and Why and How It Can Shrink

Bill Nelson

February 2024

The views expressed do not necessarily reflect those of the Bank Policy Institute’s member banks, and are not intended to be, and should not be construed as, legal advice of any kind.
How the Federal Reserve Got So Huge, and Why and How It Can Shrink

Bill Nelson
February 7, 2024

Abstract

Following the collapse of Lehman Brothers in September 2008, the Federal Reserve underwent a significant shift in how it implemented monetary policy, transitioning to an excessive-reserves framework that it had deemed too radical and rejected just months prior. This shift involved borrowing excessive reserves from banks, deviating from its traditional method of borrowing only the amount banks needed to meet reserve requirements and address clearing needs. Despite initial intentions to revert to the necessary-reserves framework, subsequent developments, including three rounds of quantitative easing, led to the permanent adoption of the excessive-reserves approach in January 2019 by the Federal Open Market Committee (FOMC). This decision was a mistake. The framework has not yielded the purported benefits, such as simpler policy implementation, and has required the Fed to be vastly larger than originally anticipated. Advocates of the excessive-reserves approach argue it aligns with the Friedman rule, but alternatives like a voluntary-reserve-requirement regime could achieve similar outcomes without the drawbacks.

Keywords: Monetary Policy; Bank Liquidity; Reserves

JEL Classifications: E52, E58, G21
Introduction

Over the six weeks following the failure of Lehman Brothers on Sept. 15, 2008, the massive expansion of credit provision by the Federal Reserve shifted the way it conducted monetary policy to a form it had considered and rejected only five months earlier in April. Commercial banks keep money on deposit at the Fed in accounts that are effectively the banks’ checking accounts. The deposits are referred to as “reserve balances”. The deposits are loans from banks to the Fed, just as ordinary retail deposits are loans from households to their banks. The Fed had been conducting policy for decades by borrowing from banks only the amount of reserve balances the banks considered necessary for payment purposes and to satisfy reserve requirements. Because the Fed had run out of options to fund its growing balance sheet, it shifted to borrowing excessive amounts of reserves from banks—that is, more than the banks needed to meet reserve requirements and satisfy payment clearing needs. In April, staff had recommended against using an excessive-reserves approach to conduct monetary policy, because “it would represent a radical departure from the basic elements of our own current framework and from those of almost every other central bank . . .” [emphasis added].

Although the Fed’s intention in late 2008 was to return to its previous method of implementing policy, over the next decade, a sequence of developments and decisions (including but not limited to three waves of quantitative easing) made it more expedient to conduct policy using an excessive-reserves approach. Staff also increased the roadblocks to returning to the necessary-reserves framework. On Jan. 29, 2019, after extensive discussion over several meetings, the Federal Open Market Committee (FOMC), the arm of the Federal Reserve that oversees monetary policy, officially decided to permanently shift to conducting policy using the excessive-reserves framework.

That decision was a mistake. The FOMC should return to conducting monetary policy using a necessary-reserves approach. The prior system had given the Fed easy, excellent control of overnight interest rates while allowing it to operate with a vastly smaller balance sheet and minimal involvement in the financial system. The excessive-reserves system has not produced its purported benefits. Policy has not been simpler to conduct, and although the federal funds rate has been stable, that’s in part because the federal funds market has essentially disappeared.

Advocates of an excessive-reserves approach have also argued that it is superior to the necessary-reserves approach because it satisfies something akin to the Friedman rule: that the inflation rate should be the negative of the Treasury bill rate, so people are indifferent between holding cash and investing in short-term, safe, interest-earning investments. The logic is that the government can offer reserves, which serve as a source of liquidity to commercial banks, for free, so it is socially optimal to provide them in abundance. As we will discuss, however, the

1 Federal Open Market Committee (2008).
2 Friedman (1969).
Friedman rule can be satisfied in a corridor system using the approach staff had recommended that the Fed adopt in its April 2008 study—a voluntary-reserve_requirement regime that pays a market rate on the self-required reserves.³

When evaluating the framework in early 2008 and again in 2018, the Fed recognized that many of the costs of the system are associated with the Federal Reserve’s necessarily large balance sheet. Indeed, as shown in Exhibit 1, the assets of the Federal Reserve System have grown from 5 percent of GDP before the Global Financial Crisis (GFC) to 28 percent at the end of 2023. The quantity of reserves that currently appears necessary to implement an excessive-reserves system is three times higher than the level the Fed judged necessary when it officially adopted the framework in January 2019, and nearly 100 times higher than estimated when the Fed considered the advisability of the framework in April 2008. That extraordinary expansion has occurred because there is a ratchet in the reserve demand process. When the Fed borrows more reserves, the system adjusts to those reserves; as a result, there is resistance when the Fed attempts to normalize its balance sheet and borrow less.⁴

One of many components of the ratchet, and the other main cost anticipated by the Fed at the outset, is that by borrowing an excessive amount of reserve balances, the Fed has caused the interbank market and the associated institutional infrastructure to evaporate. Because banks

---
⁴ Standard practice is to refer to a bank’s willingness to lend reserve balances to the Fed as its “demand” for reserves, just as a household’s willingness to lend deposits to a commercial bank is referred to as its demand for deposits. Conversely, the amount that the Fed borrows from banks in the form of deposits is referred to as the Fed’s “supply” of reserves.
have been able to earn an above-market rate by keeping elevated balances in their reserve accounts, they have less incentive to manage their liquidity tightly. As a result, the federal funds market—the market where banks with excess reserves on any given day lent to banks that were short of reserves—is essentially gone. With no interbank market, the need for each bank to maintain a high level of reserve balances gets locked in. The fact that the excessive-reserve framework leads banks to rely on deposits at the central bank as a liquid asset is why the Norges Bank decided to stop using such a system, and why some central banks are now reconsidering this approach.

A more subtle cost of the excessive-reserve system is that it imposes no natural limit on the size of the Federal Reserve’s balance sheet. Advocates of the approach believe that the Fed can borrow as much as it would like from banks without material consequences, which we will show is incorrect. By contrast, under the necessary-reserve system, if the Fed borrowed more reserves than banks wanted to keep in their accounts for clearing needs and to meet reserve requirements, the central bank would be unable to control interest rates. This perception that increased size has no cost has contributed to internal and external assumptions that the Federal Reserve’s balance sheet can be used to solve all problems, including monetizing the federal debt.

For reasons we explain in the next section, an excessive-reserves system is also called a floor system, while a necessary-reserves system is called a corridor system. This paper makes the case that the Fed should return to a necessary-reserves system and conduct policy using a corridor framework. To do so, we first present some background on the Fed’s balance sheet and the two methods of implementing policy. We then discuss why the Federal Reserve should use a corridor system.

The final section describes how the Federal Reserve can shift back to a necessary-reserves regime. If the Fed reduces the size of its balance sheet (and therefore the quantity of reserves available to the banking system), money market rates will rise above the interest rate the Fed pays banks on their deposits. This will make it increasingly expensive for banks to lend the Fed large amounts of excessive reserves. That shift will gradually overcome the ratchet effect, and the necessary amount of reserves will shrink over time. Note that once the level of reserves approaches the necessary level, the Federal Reserve will need to control large swings in reserves, as it did for most of its history. As we will discuss, it was a failure to control such swings that led to the bout of volatility in money markets in September 2019, the last time the Fed attempted to shrink.

As part of the transition back to a necessary-reserves framework, the Fed could adopt a voluntary-reserve targeting system. Each bank would set the level of reserves it wanted to hold over a maintenance period, and it would be compensated at the FOMC’s target federal funds rate for those reserves. Excess reserves would be compensated at a lower rate, and reserve shortfalls would be penalized in a way that replicates the cost of borrowing the shortfall amount from the discount window.
Such an approach separates the issue of the quantity of the reserves provided from whether the central bank implements policy using a floor or corridor system. The necessary amount is no doubt now much larger than it was before the Global Financial Crisis (GFC), because banks are both required and choose to maintain larger deposits at the Fed for liquidity management purposes. The amount banks judge necessary can be lowered through changes to bank liquidity regulations and by reducing the stigma associated with borrowing from the discount window.

If the Fed were to manage its balance sheet so that the quantity of reserves is slightly below the aggregate amount of the voluntary reserve targets, banks would need to turn to the discount window at the end of the maintenance period. The need to borrow, and the modest upward bias to money market rates, would offer an incentive for banks to economize on their use of reserves as a liquidity management tool, while the increased borrowing—especially as an intended component of the regime—would help lessen stigma. The resulting small average spread between money market rates and the rate paid on voluntary reserves would capture the social costs associated with operating the central bank with a large balance sheet.

**Background**

Understanding monetary policy implementation requires understanding central bank balance sheets. Exhibit 2 presents a stylized version of the Federal Reserve’s balance sheet.

**Exhibit 2**

Simplified Federal Reserve Balance Sheet

The Federal Reserve’s assets consist primarily of securities—Treasury securities and, during some periods, agency debt and agency mortgage-backed securities. The Fed is required to
purchase these securities in the open market, and they are collectively referred to as open market operation or “OMO” securities. The other main Fed assets are loans to banks (discount window loans); repurchase agreements against OMO securities; and, during crises, loans to nonbanks. We will follow the Fed’s convention of referring to repos and reverse repos from the perspective of the Fed’s counterparty. The Fed is therefore actually engaged in a reverse repo (lending cash against OMO collateral) when it engages in a “repo,” which is why repos are listed as an asset, not a liability.

The liabilities the Fed uses to fund its asset holdings consist primarily of currency, reserve balances, the Treasury’s deposit at the Fed (the “Treasu ry General Account” or “TGA”) and reverse repos. Currency is a zero-interest loan from the public (especially foreigners) to the Federal Reserve. Until the Lehman default in September 2008, the Fed funded itself almost entirely with currency. However, after Lehman, the Fed’s borrowing needs exceeded the amount of currency, so it began to borrow heavily from depository institutions: commercial banks, thrifts, and credit unions (“banks”). Banks offer funding to the Fed in the form of deposits at their Federal Reserve Bank. Those deposits are also known as “reserve balances.” At times, the Fed has needed to also borrow from money market mutual funds, GSEs, foreign official institutions and primary dealers (the broker-dealers authorized to be Fed counterparties) using overnight reverse repos collateralized by OMO securities. In 2013, the Fed began borrowing overnight reverse repos (ON RRP s) at a standing facility as a tool to support monetary policy. Reverse repos had usually been zero before the GFC. In 2021–22, they were over $2 trillion but they have fallen substantially since, and as of Jan. 25, 2024, they were $558 billion.

The Federal Reserve’s book equity (using its own accounting conventions) is close to zero. In 2015, Congress expropriated the Fed’s equity to help pay for the Fixing America’s Surface Transportation Act and limited the equity going forward to $10 billion. In 2021, Congress limited the Fed’s equity to $6,785 million.

The Federal Reserve conducts monetary policy by setting its policy rates and adjusting its balance sheet so that the federal funds rate (the interest rate on unsecured loans between banks or between banks and GSEs) trades near the FOMC’s target or within its target range. The Fed’s policy rates are the interest rate at which it lends to banks (the “primary credit rate” a/k/a the “discount rate”) and the interest rate it pays on deposits (the “interest on reserve balance” or “IORB” rate). The Federal Reserve adjusts its balance sheet by engaging in repos or reverse repos or asset purchases or sales to expand and contract its assets.

---

5 See Nelson (2021) for more information.
6 The Fed does not follow GAAP accounting rules; it follows its own. See Board of Governors of the Federal Reserve System (2023a).
7 If Congress had taken the equity but allowed the Fed to build it back up by retaining earnings rather than remitting them to Treasury, the expropriation would not have scored as a source of revenue in the budget projection.
8 The Fed switched from targeting a level for the fed funds rate to targeting a range on Dec. 16, 2008.
9 The Fed set reserve requirements to zero in March 2020. Originally, the relevant policy rate was the interest on excess reserves (IOER) rate, but we will mostly refer to the rate as the IORB rate.
Because currency, the TGA and reverse repos are unaffected by these transactions, and balance sheets must balance, by adjusting its assets the Fed also adjusts reserve balances. Setting aside the interaction between reserves and ON RRPs, the aggregate level of reserve balances—the amount the Fed needs to borrow from banks—is immutable: unchanged by transactions between banks or other financial institutions. At the same time, each bank can freely choose the reserve balance it wants. The federal funds rate, as well as other interest rates, adjusts so that banks each choose to keep deposit balances at the Fed that add up to the aggregate amount of reserve balances necessary to fund fully the Fed’s balance sheet. That’s how monetary policy works.

Banks choose to keep overnight deposits at the Fed—reserve balances—for various reasons. Until March 2020, banks were required to maintain reserves, on average over a two-week period, equal to a fixed percentage of their deposits; that is, banks were subject to a reserve requirement. In addition, a bank’s deposit at the Fed is its checking account, and banks maintain positive amounts in their accounts to support their payments activity. A deposit at the Fed is also the ideal source of contingency funding—it is perfectly safe and immediately available, and its value does not change. Banks currently keep a large amount of reserves for liquidity management purposes.

For half a century, the relationship between reserve balances and the federal funds rate has been conceptualized using a model William Poole developed in 1968 while he was an economist at the Federal Reserve Board. The model is based on the behavior of a bank that must decide how much to borrow in the federal funds market at mid-day without knowing what transactions will hit its account in the afternoon. Note that the model describes the behavior of a bank over the course of one day, not over time.

Although not exactly how Poole presented it, the model as currently used is shown in Exhibit 3. The orange line is the quantity of reserves borrowed by the central bank each day, which is fixed following the morning repo operation, to adjust the size of the balance sheet. The purple line is the “demand” by banks for reserve balances, which describes what federal funds rate will correspond to any particular supply. To the left, the curve flattens out at the discount rate, because a bank should not be willing to borrow in the funds market for more than it can borrow from the Fed. To the right, the curve flattens out at the IORB rate, because a bank should not be willing to lend in the funds market for less than it can receive by simply keeping its funds on deposit at the Fed.

---

10 Poole (1968).
In the middle of Exhibit 3, there is a relatively steep portion of the line connecting the two flat regions. The relationship corresponded well with the behavior of the federal funds rate during the day when the Federal Reserve sought to supply the quantity of reserves demanded by banks and paid zero interest on deposits. Actually, the correspondence really only worked well on the last day of the two-week maintenance period over which banks had to meet their reserve requirements on average. For other days, the federal funds market generally cleared at the FOMC’s target rate, irrespective of modest changes in the quantity of reserve supplied or demanded, because banks knew that they could make up any excess holding or shortfall by adjusting their balances later in the maintenance period.

At the end of the maintenance period, if the Fed inadvertently created more reserves than banks wanted, the federal funds rate dropped sharply at the end of the day, because all market participants sought to shed the reserves on which they would earn zero. If the Fed inadvertently created insufficient reserves, the federal funds rate would rise until it reached the discount rate. Banks would then borrow from the discount window, expanding the Fed’s balance sheet and creating the necessary reserves. In fact, as we will discuss, banks were and are reluctant to use the discount window, so the federal funds rate would rise somewhat above the discount rate before banks would borrow.

The model is well suited for describing corridor and floor monetary policy implementation frameworks. The federal funds rate is determined within a corridor created by the discount rate, which is the ceiling, and the deposits rate, which is the floor. As shown in Exhibit 3, if the central bank is operating using a corridor system, it sets a target for the middle of the corridor and sets the supply of reserve balances in the middle of the steep section of the demand curve.
If the central bank is using a floor system, it sets the target rate at the deposit rate and sets the supply of reserves well to the right of the steep slope.

It is a convenient fiction to say that the Fed operated policy using a corridor system before the GFC. From the mid-1960s to 2002, the discount rate was below-market. The Fed controlled rates using the upward supply of reserves induced by the reluctance to borrow at the discount window, changes in reserve supply through open market operations, and the downward sloping demand for reserves. From 2003 (when the Fed changed the discount rate to an above-market rate) to October 2008, the Fed had a ceiling, but the floor was zero because the Fed was not allowed to pay interest on reserves.

On Oct. 1, 2008, the Federal Reserve began paying interest on reserve balances. The Financial Services Regulatory Relief Act of 2006 authorized the Federal Reserve Banks to pay interest on bank balances, effective Oct. 1, 2011.11 The start date was advanced to Oct. 1, 2008, by the Emergency Economic Stabilization Act of 2008.12 The Fed had sought the authority to pay interest on required reserves, not excess ones.13 Paying interest on required reserves eliminates the tax on banks caused by forcing them to hold unremunerated required reserves: forced zero-interest loans to the government. Moreover, doing so eliminated the incentive for banks to minimize their reserve requirements by sweeping deposits at the end of the day into money fund accounts. Sweep accounts had become so common that the level of required reserves was falling below the level necessary to implement monetary policy.

However, Congress accelerated the date for the Fed to start paying interest on reserves so that the Fed could pay interest on excess reserves as well as, incidentally, required reserves. It was necessary for the Fed to pay interest on excess reserves to conduct monetary policy once the Fed began offering more reserve balances than banks demanded as a consequence of the post-Lehman expansion of the balance sheet. By creating excessive excess reserves, the Fed switched from a corridor to a floor framework for monetary policy implementation. And as can be seen in Exhibit 4, when the Fed began to oversupply reserves, the federal funds rate fell well below the interest on excess reserves (IOER) rate. At the time, the IOER rate was 1 percent and the FOMC’s target for the fed funds rate was 2 percent, but the funds rate fell essentially to zero. The IOER rate’s failure to create a floor on the funds rate was moot by December 2008, because the Fed set a target range for the funds rate of 0 to 25 basis points. But the failure cast doubt on the effectiveness of the IOER rate to put a floor on the funds rate. This would become important five years later, when the FOMC began to contemplate raising the funds rate above zero.

---

Although the Poole model describes the intraday behavior of the fed funds market well, it has proven a poor guide for how the demand for reserves behaves when the Fed pays interest on reserve balances and borrows massively more from banks than necessary. As Poole (1968) stated:

The model presented here concentrates on these very short-run adjustments. However, it is obvious that the bank must make further adjustments if it experiences persistent reserve drains or accretions. (p. 770)

As shown by Afonso and colleagues (2022, including John Williams, the president of the New York Fed), the minimum level of reserve balances required by banks has shifted to the right over time.\(^4\)

Indeed, as reported in Exhibit 5, the Fed’s estimate of the quantity of reserve balances required to implement a floor system has risen dramatically over time, in line with the tremendous growth in reserve balances. In the original staff memo from April 2008 on methods to implement policy using interest on reserves, staff estimated that $35 billion would be necessary to implement policy in a floor system. In March 2016, the New York Fed raised its assumption to $100 billion.\(^5\) In March 2018, it raised it again to $600 billion.\(^6\) At the December 2018 FOMC meeting, at the cusp of the Fed deciding to adopt a floor system permanently, staff judged that

\(^4\)Afonso and colleagues (2022)
$1 trillion would be needed, $800 to supply the reserves required by banks and a $200 billion buffer to absorb volatility in the supply caused by changes in other balance sheet items, particularly the TGA.¹⁷ By September 2019, it had more than doubled this to $1.3 trillion.¹⁸ In the New York Fed’s forecast of the Fed’s balance sheet published in May 2022, the estimate essentially doubled again to $2.3 trillion. Responses to the Fed’s May 2023 Senior Financial Officer Survey found that a notable fraction of banks had raised their target level of reserves substantially since late 2022.¹⁹

Nelson (2018a, 2022) argues that the growth is the result of bank management and bank examiners becoming accustomed to addressing bank liquidity needs with subsidized reserve balances. For example, in 2010, concern about the negative consequences of holding excess reserves led Norges Bank (the central bank of Norway) to switch from a system with abundant reserves to one with more scarce ones. When seeking comment on their decision, they noted:

> When Norges Bank keeps reserves relatively high for a period, it appears that banks gradually adjust to this level. . . . With ever increasing reserves in the banking system, there is a risk that Norges Bank assumes functions that should be left to the market. It is not Norges Bank’s role to provide funding for banks. . . . If a bank has a deficit of reserves towards the end of the day, banks must be able to deal with this by trading in the interbank market.²⁰

Alternatively, Acharya, Chauhan, Rajan, and Steffen (2023) point to banks’ changing their balance sheets in ways that require the added liquidity from reserve balances, such as offering more lines of credit. Similarly, Lopez-Salido and Vissing-Jorgensen (2023) argue that reserve demand increases with the level of bank deposits. Consequently, the current conceptualization of reserve demand and supply, and the behavior of money market rates, is not the Poole model with its stable demand curve and fixed steep section. Instead, similar to the relationship between short-run and long-run production functions familiar to microeconomists, the steep part of the demand curve moves gradually up and down, in reaction to the available supply of reserve balances and configuration of interest rates.

---

¹⁸ For further discussion of the Fed’s unfortunate decision to hold back its forecast of its balance sheet and income in March 2019, see Nelson (2019a).
¹⁹ Board of Governors of the Federal Reserve System (2023b).
²⁰ Norges Bank (2010).
If it still seems helpful to use the Poole model as an intuitive source, the relationship between the level of reserve balances and the federal funds rate appears to look like Exhibit 6. As the Fed offers more reserves, the steep part of the demand curve shifts right over time. Moreover, rather than being flat at the IORB rate, the demand curve slopes gradually down—the more reserves the Fed borrows from banks, the more it has to pay for them, relative to market rates.
How the Federal Reserve Ended Up Using a Floor System

The Fed adopted a floor system through a sequence of accidents, expediency and hard-to-reverse choices. During the earlier stages of the GFC, the Fed reduced its holdings of short-term Treasury securities to make room for the increased lending without having to borrow more from commercial banks in the form of reserve balances. Moreover, starting on Sept. 17, 2008, the U.S. Treasury initiated the “Supplementary Financing Program” at the request of the Federal Reserve. The Treasury issued extra Treasury bills and then deposited the proceeds with the Fed, offering more financing for the Fed’s balance sheet. Nevertheless, by the end of October 2008, with the Fed’s alternatives exhausted, it began to borrow more from banks than the banks needed to keep on deposit, entering an excessive-reserves framework.

The FOMC massively expanded the Fed’s portfolio of securities—and its borrowing from banks to finance its asset holdings—with several waves of large-scale asset purchase programs, known as quantitative easings. On Nov. 25, 2008, the Fed announced that it would purchase up to $100 billion in direct GSE obligations and $600 billion in agency MBS. The purchases were made “to reduce the cost and increase the availability of credit for the purchases of houses, which in turn should support housing markets and foster improved conditions in financial markets more generally.” A few months later, on March 18, 2009, the FOMC announced that it would increase its agency debt purchases to $200 billion and its agency MBS purchases to $1.25 trillion, and that it would also purchase $300 billion in longer-term Treasury securities.

When the LSAPs first began, the Fed planned on letting the MBS and agency debt roll off as they matured. But in August 2010, when MBS repayments picked up in response to a drop in mortgage rates, the Committee decided to reinvest the principal payments into longer-term Treasuries. On Nov. 3, 2010, “[t]o promote a stronger pace of economic recovery and to help ensure that inflation, over time, is at levels consistent with its mandate,” the Committee announced that it would purchase another $600 billion in longer-term Treasury securities.

On June 12, 2011, the Committee announced “Exit Strategy Principles” to govern how it would normalize its balance sheet. Sometime before raising the funds rate, the Committee would start allowing its securities holdings to roll off as they matured, and sometime after first raising the funds rate, it would begin to sell its holdings of agency debt and agency MBS. The Committee expected to eliminate those securities holdings over three to five years. It anticipated that reserve balances would return to the “smallest level consistent with the efficient implementation of monetary policy” in two to three years. The staff forecast of the Fed’s balance sheet indicated that the balance sheet would be normalized in November 2015.

---

23 Federal Open Market Committee (2009).
24 Federal Open Market Committee (2010a).
25 Federal Open Market Committee (2010b).
26 Federal Open Market Committee (2011a).
when reserve balances reached $25 billion, consistent with the pre-GFC necessary-reserves approach.\textsuperscript{27}

It was QE 3, the “flow-based” asset purchase program (a/k/a “QE-infinity”) that set the Fed on a path to adopting an excessive-reserves system permanently. On Sept. 13, 2012, the Committee announced that it would start buying $40 billion in agency MBS each month, in addition to the $45 billion in longer-term Treasury securities it was buying as part of the Maturity Extension Program (MEP).\textsuperscript{28} The purchases of longer-term Treasury securities under the MEP were being offset by sales of shorter-term securities. On Dec. 12, 2012, the Committee announced that it would continue buying $40 billion in agency MBS and $45 billion in Treasury securities each month, but no longer sell an offsetting amount of shorter-term securities (which were running low) for a total of $85 billion in monthly purchases. The Committee announced that it would continue the purchases until the “outlook for the labor market . . . improve[d] substantially.”\textsuperscript{29}

As Nelson (2018b) describes, when the program was first presented to the Committee (but not as the recommended choice) in the July/August 2012 meeting, the staff projected it would continue through the third quarter of 2013 and the purchases would total $1 trillion.\textsuperscript{30} But when the program was presented as the proposed policy choice in the October and December 2012 meetings, it had somehow become smaller—the staff projected that it would end in June 2013 and total $750 billion.\textsuperscript{31} Committee members were skeptical, because the staff forecast for the unemployment rate and projected gains in employment in mid-2013 showed no improvement. For example, then-Governor Jay Powell asked:

How is this not a $1 to $2 trillion LSAP [large scale asset purchase program]? Where is the improvement in labor markets?\textsuperscript{32}

In Powell’s comments in the December 2012 Survey of Economic Projections, he expected that QE3 would begin to wind down in mid-2013 as the staff projected, not because of an improvement in labor markets, but rather because the program would have proven ineffective.\textsuperscript{33} In the event, QE3 continued until October 2014, 15 months later than originally projected by the staff, and totaled $1.75 trillion in MBS and Treasury securities, more than double the amount originally projected.

One of the Committee’s concerns about QE3 was that the staff projected that the Fed would have losses when it sold its agency MBS. The staff briefed the Committee on losses in both September and December 2012. Chairman Bernanke and FOMC Vice Chairman Dudley (president of the New York Fed) argued that the Fed could avoid losses by allowing securities to roll off as they matured, rather than selling them and so be able to continue remitting profits to

\textsuperscript{27} Board of Governors of the Federal Reserve System (2011).
\textsuperscript{28} Federal Open Market Committee (2012a).
\textsuperscript{29} Federal Open Market Committee (2012b).
\textsuperscript{30} Board of Governors of the Federal Reserve System (2012a).
\textsuperscript{31} Board of Governors of the Federal Reserve System (2012b) and (2012c).
\textsuperscript{32} Federal Open Market Committee (2012c).
\textsuperscript{33} Federal Open Market Committee (2012d), p. 27, respondent 15.
Treasury. For example, in December 2012, Bernanke stated, “We also have some ability to affect the pattern of remittances, and, indeed, as was pointed out, by not selling securities, we can avoid long periods of no remittances.”³⁴ In a foreshadowing, Dudley noted that doing so could be consistent with implementing policy using an excessive-reserves approach.

. . . a lot of this also depends on what monetary framework you are actually going back to. And the presumption of the staff memo is that we are going back to a corridor system. But you might decide, as you go through this, that maybe IOER works pretty well. And you might actually want to go back to a floor system that would allow you quite a bit more discretion in terms of how your exit actually works. And that would feed into a lot of these projections.³⁵

While the program was under way, on Sept. 16, 2014, the Committee officially changed its “normalization principles and plans”³⁶ to reflect the new plan of not selling securities. The new principles stated that the Committee no longer intended to sell agency debt or MBS but would instead simply allow the securities to mature without reinvestment. Reserve balances would be reduced to “no more . . . than necessary to implement monetary policy efficiently and effectively. . . .” The Committee no longer indicated that it would implement policy using the “smallest” level of reserves possible. In the material given to the FOMC at the September 2014 meeting, the “normalized” level of reserve balances was now projected to be $100 billion.³⁷

With interest rates at zero and the size of the balance sheet unrestricted, Fed staff made decisions about the Fed’s other liabilities that made it much harder to return to a necessary-reserves framework. Before the GFC, the U.S. Treasury had kept most of its cash in deposits in the banking system, normally leaving about only a $5 billion balance at the Federal Reserve. Treasury deposits at commercial banks were collateralized and operated under the Treasury, Tax, and Loan (TT&L) program. Exhibit 7 shows what happened in October 2008, when there was a “change in the Federal Reserve-Treasury agreement” and Treasury increased its balance substantially.³⁸

Because the Fed was becoming so massive and borrowing so heavily from banks, it had to pay banks more for reserve balances than the Treasury earned in interest on its deposits. It therefore became cost-effective for the Treasury to keep its deposit at the Fed instead. Even though the Treasury did not earn interest on the TGA, taxpayers were better off on net when reserve balances (with their above-market rate) were reduced.

Over time, the TGA continued to grow and become more volatile with especially large swings around debt-ceiling debacles. Because the TGA is a Fed liability, when it increases, the Fed does not need as much funding from banks, so reserve balances decline dollar-for-dollar; and when

---

³⁴ Federal Open Market Committee (2012e).
³⁵ Federal Open Market Committee (2012c).
³⁶ Federal Open Market Committee (2014).
the TGA shrinks, reserve balances rise. If the Fed were to return to a necessary-reserves framework, the TGA’s volatility would be a challenge to offset, keeping the level of reserve balances relatively constant. We will return to this issue when we discuss how the Fed could go about returning to a small balance sheet.

![Exhibit 7: U.S. Treasury General Account](image)

Another Fed liability item that has grown significantly since the GFC is “Reverse Repurchase Agreements: Foreign Official and International Accounts,” a/k/a the “Foreign Repo Pool.” Foreign official and international holders of accounts at the Fed invest in overnight reverse repurchase agreements that offer the Fed cash and receive securities held in the Fed’s securities portfolio. They are paid an interest rate tied to comparable market-based interest rates. Foreign official and international institutions put funds into the foreign repo pool because the reverse repos are a safe and highly liquid dollar investment. The Fed indicates that it offers banking and financial services to “about 250 central banks, governments and international official institutions,” but the list of institutions is not made public.

Whereas before the GFC, the foreign repo pool was limited to about $25 billion, it has risen steadily and is now in the hundreds of billions of dollars, with large, rapid swings that result in corresponding changes in reserve balances (Exhibit 8). In Markets Group of the Federal Reserve Bank of New York (2017), the Fed described the reasons for the increase:

> The rise in balances in recent years reflects central banks’ preferences to maintain robust dollar liquidity buffers, the reduced availability of alternative investment options
with private counterparties, and the New York Fed’s removal over time of constraints on customers’ ability to vary the size of their investments. (p. 30)

The Fed likely reduced the constraints it had been imposing on how customers could “vary the size of their investments,” because doing so was unnecessary under the excessive-reserves framework.

Note that there is no record that the decisions to let the TGA increase or to expand the foreign reverse repo pool were discussed with the FOMC, even though they became important reasons why the FOMC could not return to a necessary-reserves framework. In 2016, for example, staff told the FOMC that much work would need to be done if the Committee wanted to control TGA volatility, that banks probably wouldn’t want the Treasury’s deposits, and that any changes would need to be phased in. Staff pointed out that the Committee might see a “... public benefit to supporting the Treasury’s objective to hold a buffer to meet its obligations.” Staff concluded: “[i]n contrast, a floor framework can accommodate large and volatile autonomous factors.”

As part of conducting policy with a massive quantity of reserve balances, the Fed needed to substantially broaden the set of institutions from which it borrowed money. As it turned out, the federal funds rate was not pinned down by the IORB rate when the Fed sought to borrow

---

more and more from banks. Instead, the more the Fed borrowed, the more the fed funds rate (and other money market rates) fell below the IORB rate. That is, the price the Fed paid to borrow from banks, measured relative to comparable market rates, rose the more the Fed sought to borrow.

To ensure that the Committee would, in fact, be able to raise the fed funds rate when it wanted to, in 2013 the Fed opened a standing facility where nonbanks—primarily money market mutual funds, but also GSEs and primary dealers—could lend to the Fed through overnight reverse repurchase agreements: the ON RRP facility. When the FOMC began to raise the funds rate from the zero lower bound starting in December 2014, use of the facility increased, averaging about $150 billion until declining essentially to zero in 2018.

The facility rose again, and much higher, following the COVID crisis. The facility had remained near zero during lockdown, even though the Fed’s massive asset purchases increased reserve balances from $1¾ trillion to $3½ trillion in two months. When the pandemic struck, the Fed had excluded reserve balances from the calculation of bank capital requirements, so the expansion was largely without cost to banks, especially because they were being inundated with deposits. However, at the end of March 2021, the temporary exclusion ended, and banks once again had to fund their loans to the Fed in part with capital. Instead, it became cheaper for money funds, which have no capital requirements, to finance the Fed at the margin. The ON RRP facility rose rapidly, eventually peaking at $2.5 trillion by the end of 2022.

This episode illustrates how the shift to a floor system was partly facilitated by the Committee becoming comfortable with a giant balance sheet and expanded counterparties. Although Committee members including Governors Powell, Stein and Tarullo expressed serious misgivings in 2012 about expanding the Fed’s set of counterparties to include money market mutual funds, in May 2023 Chair Powell stated that the ON RRP facility (which was then more than 10 times its peak level in 2014) was simply doing the job it was designed to do.

Over time, the Fed’s plan shifted from returning to a corridor system to remaining in a floor system. In part, that shift reflected a change in which approach was considered the default. Typically, the stated reason to demur was that the Committee didn’t need to decide yet and was gaining valuable information. Bill Dudley, president of the New York Fed, gave that as the reason at the June 2011 FOMC meeting:

Now, in all of this, I’m not saying I favor a floor system. I’m not arguing for a floor system over a corridor system. All I am arguing is that it is bad policy to rule out a floor

---

40 The Fed opened the ON RRP facility in September 2013 for experimentation and officially adopted it as a tool for normalizing policy in September 2014.
41 When Powell was asked in the press conference following the May 2023 FOMC meeting if the ON RRP facility was making the deposit outflows from banks worse, he responded: “[The facility] is really there to, to help us keep rates where they’re supposed to be, and it’s, it’s serving that purpose very well.” Federal Open Market Committee (2023), p. 20.
system at the current time given the chance to learn a lot about how a floor system operates in practice.\textsuperscript{42}

And six years later, that was how Janet Yellen, chair of the FOMC, explained why they had not picked a destination for normalization, even when they were beginning to shrink the balance sheet in June 2017:

I can’t tell you what the longer-run normal level of reserve balances will be because that will depend on the Committee’s eventual decisions about how to implement monetary policy most efficiently and effectively in the longer run, as well as a number of as-yet unknown elements, including the banking system’s future demand for reserves and various factors that may affect the daily supply of reserves. . . . Decisions about the appropriate long-run framework do not need to be made for quite some time, and our future deliberations will benefit from the experience we will gain during the normalization process.\textsuperscript{43}

Borio (2023) argues that it is hard to believe that it would be any less easy to operate a corridor system now than it once was, noting that most central banks still use a corridor system. He observes that “[s]ome of the concerns may simply reflect the loss of institutional memory within central banks.”\textsuperscript{44} Consistent with Borio’s insight, in a memo to the FOMC as background for the discussion of the implementation framework at the November 2018 meeting, staff described the excessive-reserve approach as simple to implement while the pre-GFC approach required “a high degree of complexity”.\textsuperscript{45}

**How the Fed Conducted Policy in the Necessary-Reserves Framework**

As Borio observed, there is a widespread misconception that conducting policy in a necessary-reserves framework was difficult and complicated. That misunderstanding has arisen in part because it seems more difficult and complicated to supply reserves in the middle of the steep part of the demand curve than somewhere on the long flat section to the right. In addition, the floor system is somewhat easier to understand; the 15 years since the Fed operated under its previous regime has also significantly reduced familiarity with it.

However, implementation was relatively simple and robust. Each morning, staff at the New York Fed and Board would project the autonomous factors that would shift the demand for and supply of reserves, such as tax payments. FRBNY would speak with the funding officers at large banks and ask them how much in reserve balances they anticipated needing that day. Putting the information together, staff would decide whether a temporary operation was needed to adjust the supply of reserves. If necessary, the Desk would conduct repo operations with the primary dealers. The operations were relatively small and unimportant in primary dealers’

\textsuperscript{42} Federal Open Market Committee (2011b), p. 37.
\textsuperscript{43} Federal Open Market Committee (2017).
\textsuperscript{44} Borio (2023), pp. 8–9.
\textsuperscript{45} Federal Reserve System (2018)
businesses. (Primary dealers do not have accounts at the Federal Reserve and therefore do not hold reserve balances.)

The Fed was thus engaged in fairly small transactions with a limited set of counterparties to influence conditions in the interbank market, where it did not intervene directly, apart from the rare discount-window loan. The autonomous factors were well understood and accurately projected. If there was a mismatch between demand and supply, it did not move the funds rate by much, except on the last day of the maintenance period. As a result, the Desk was able to keep the funds rate near the FOMC’s target with a small footprint in financial markets.

Policy implementation was also made easier because when the Fed signaled where it wanted the federal funds rate to be, the funds rate would move to that new level. Borio (2023) notes:

> Signalling did the heavy lifting in steering the overnight target. Through signals, the central bank would tell banks where it wanted the overnight rate to be. In the meantime, light-touch liquidity management operations worked only in the background. They would keep the market for bank reserves in balance and ensured that the demand for settlement balances “did not get in the way”.

The effectiveness of what are sometimes called “open-mouth operations” is likely due in large part to the market’s awareness that the Fed could achieve its target by adjusting the supply of reserve balances and changing the discount rate.

**Why the Fed Should Stop Using an Excessive-Reserves Framework**

The benefits of an excessive-reserves framework have been modest at best and the costs have been high. Moreover, as we discuss in the next section, the Fed can shift back to a necessary-reserves framework (albeit one adjusted to fit current circumstances) without generating turmoil in money markets.

**The Scant Benefits . . .**

In multiple studies, Fed staff and others concluded that a central bank can achieve good interest rate control using various approaches, including both a necessary-reserves framework and an excessive-reserves framework. Rather than interest rate control, the Committee saw two principal advantages in a floor system when it evaluated framework alternatives in 2018 just before officially adopting a floor system in January 2019. First, it judged that the system would be simpler. Because the Committee envisioned the supply of reserve balances would always be on the flat part of the demand curve, variation in supply as well as demand shifts left

---

46 Borio (2023), p. 3.

47 See the staff briefing on the options in the transcript to the November 2016 FOMC meeting, Federal Open Market Committee (2016), pp. 7–10. See also, for example, the 2016 staff memo to the FOMC, “The Foreign Experience with Monetary Policy Implementation” (Federal Reserve System [2016b], p. 1), which states: “Nearly all advanced economy central banks have implemented monetary policy with short-term interest rate targets; other aspects of their policy implementation regimes vary widely, largely due to central banks’ particular environments and institutional history. Even so, all of the central banks studied seem to have control over the level of short-term interest rates and have been able to transmit policy effectively to longer dated yields.”
or right should not materially move the federal funds rate. Irig, Senyuz, and Weinbach (2020) state: “[a]t that time, the Fed will be in position to conduct OMOs over time solely to accommodate trend growth in autonomous factors to maintain an ample level of reserves. . . .”\textsuperscript{48} Similarly, in Federal Reserve System (2016a), staff asserted that “[i]n a floor system, there is generally little need for detailed information on the demand curve for reserves, and the regime is robust to shifts in demand and shifts in supply. . . .”\textsuperscript{49}

The second perceived advantage was that if the Committee used a floor system in normal times, it would not need to change to a floor system if a crisis required it to expand its balance sheet. In the November 2016 FOMC discussion of operating regimes, for instance, Committee Vice Chair and President of the New York Fed William Dudley listed as an advantage of a floor system, “If you had a corridor system, reserves added by QE and special liquidity programs would have to be drained in order to maintain monetary control, and we saw in 2008 that this can be difficult to do.”\textsuperscript{50}

Regarding the first purported advantage, implementation has been far from set-it-and-forget-it, both because of problems when reserve balances turned out to be too low and when they ended up too high. It has been impossible to predict the amount of reserve balances below which interest rates would rise. In September 2019, the misidentification of the necessary supply of reserves, coupled with a failure to conduct an OMO to offset a large drop in reserves, led to a severe bout of volatility in repo markets. This necessitated a massive injection of reserves using repos. The September 2019 episode, and how to avoid repeating it, is discussed in the next section.

Moreover, as we noted, to maintain control of interest rates as reserve balances rose in response to the COVID asset-purchase program, the Fed needed to reactivate the ON RRP facility. At the peak, it consistently engaged in over $2\textsuperscript{11/4} trillion in reverse repos with money funds. Taken together, as shown by Exhibit 9, temporary open operations by the New York Fed (the standard measure of monetary policy fine-tuning) have been much larger and more volatile in the years since the Fed began conducting policy in a floor system than in the prior decade, when operations were frequent but small.\textsuperscript{51}

\textsuperscript{48} Irig et al. (2020), pp. 24–25.
\textsuperscript{49} Federal Reserve System (2016a), p. 6.
\textsuperscript{50} Federal Open Market Committee (2016).
\textsuperscript{51} The Federal Reserve defines two types of open market operations, permanent and temporary: “The Federal Reserve conducts OMOs in domestic markets. OMOs can be permanent, including the outright purchase and sale of Treasury securities, government-sponsored enterprise (GSE) debt securities, and federal agency and GSE MBS; or temporary, including the purchase of these securities under agreements to resell, and the sale of these securities under agreements to repurchase.” Board of Governors of the Federal Reserve System (2019).
Regarding the second objective, the framework has supported good interest rate control through periods of stress, including the end of the GFC, the COVID period, and the banking turmoil in spring 2023 (Exhibit 10A). However, as noted, achieving interest rate control in the mid-2010s and during COVID when the Fed’s balance sheet was especially large required the addition of the ON RRP facility. In addition, for the most part, at times the FOMC’s target was effectively at zero, which likely supported the rate. And finally, as we will discuss, because the Fed was massively oversupplying reserve balances, the fed funds market stopped being a place where banks lent and borrowed from each other for liquidity purposes. Instead, the market consist of transactions between Federal Home Loan Banks and the U.S. branches of foreign banking organizations at rates that varied little day to day.\(^52\) For example, participants in a January 2019 symposium on money market developments noted that the fed funds rate had not risen over the prior year-end, even though repo rates had risen over 50 basis points. The participants “saw that stability as reflecting the idiosyncratic nature of transactions in the fed funds market currently with the result that the fed funds rate was now a poor metric for money market conditions.”\(^53\) Similarly, over the 2023 turn, repo rates spiked and the volume of transactions in the federal funds market fell by one-third, but the effective federal funds rate was unchanged. Indeed, the reduction in variability, while still significant, is less pronounced for the GC-repo rate than for the federal funds rate (Exhibit 10B).

\(^{52}\) Afonso and colleagues (2023).
\(^{53}\) Nelson (2019b).
Furthermore, it is by no means clear that the Fed would have been unable to maintain control during crisis periods if it had been operating policy with a corridor system, because that system becomes a floor system when the central bank increases reserve balances substantially. That is exactly what happened to the ECB when it increased its balance sheet after Lehman in 2009; during the European banking crisis in 2012; and in 2015, when it increased its purchases of government debt (Exhibit 11). As reserve balances rose, the Eurosystem shifted smoothly from conducting policy with a corridor system to conducting it with a floor system. For example, the ECB’s August 2009 *Monthly Bulletin* describes how the unsecured interbank rate declined toward the deposit rate as the ECB added excess liquidity in the form of longer-term loans to banks.\(^{54}\) Similarly, the June and September 2015 *Economic Bulletins* describe how money market rates edged down to the central bank’s deposit rate as it continued to borrow more from banks to fund its longer-term refinancing operations and purchases of government bonds.\(^{55}\) Taking a longer view, a 2023 Bank of Finland article describes how the continued provision of excess reserves transitioned the ECB from a corridor to a floor system, with no report of any difficulties.\(^{56}\)

\(^{54}\) European Central Bank (2009).

\(^{55}\) See the boxes on liquidity conditions and monetary operations in the two economic bulletins: European Central Bank (2015a) and (2015b).

\(^{56}\) Herrala and Tötterman (2023).
Another benefit cited for a floor system is that it satisfied a condition like the Friedman rule. In particular, because the approach leaves money-market rates close to the interest rate the Fed pays on deposits, liquidity is essentially free. In a 2023 speech, “Ample Reserves and the Friedman Rule,” Lorie Logan, president of the Dallas Fed, observed:

That is, in a floor system, banks’ demand for central bank reserves is satisfied at something close to market interest rates. Market rates are not materially higher than the remuneration on the marginal dollar of reserves, and correspondingly there is no incentive for banks to economize on reserves, the most liquid asset in the financial system. As a result, floor systems satisfy a version of the Friedman rule. The opportunity cost to banks of holding reserves is approximately equal to the central bank’s cost of supplying reserves, which most analysts view as small.

However, the Friedman rule can be satisfied by implementing policy in a corridor system using the method recommended by staff in the initial April 2008 study of the Fed’s options. Specifically, staff recommended that policy be implemented using a voluntary reserves targeting regime. Each bank picks a level of reserve balances it wants to hold over each period between FOMC meetings. Reserves that the bank ends up holding on average within a band around the voluntary target could be remunerated at the FOMC’s target for the federal funds rate. If the bank ends up holding excess reserves on average, the interest rate on the reserves above the target band would be substantially below the FOMC’s target rate (50 or 100 basis points, for example). If the bank ends up with reserve balances that are on average below the band, its remuneration would be adjusted to mimic the cost of borrowing the needed reserves from the discount window.

Note that meeting the voluntary reserve target would be optional, apart from the financial consequences. Mandatory required reserves are not eligible to satisfy liquidity requirements, but optional required reserves are eligible. Because the Fed’s reserve requirements were mandatory, required reserves were not counted as a source of liquidity before reserve requirements were set to zero.

If a bank wished to hold a high level of reserve balances, it could set its voluntary target to a high level, and it would receive a market rate on those reserves. For liquidity purposes, reserve balances would satisfy the Friedman rule. On any particular day, banks with extra reserves would have an incentive to trade in the fed funds market with banks that were short of reserves, so the fed funds market would revive.

---

57 In Milton Friedman and Economic Debate in the United States, 1932–1972, Volume 1, Ed Nelson (2020) argues that the term “Friedman rule” should only be applied to Friedman’s constant money growth rule. Although Friedman derived (along with others) the deflation/satiation rule for the quantity of money, he did not endorse it.
58 Logan (2023).
This approach would allow the policy implementation system to be separated from that of the use of reserve balances to satisfy liquidity requirements, and therefore the size of the Fed’s balance sheet. The Fed cannot both require banks to hold large reserve balances for liquidity purposes and also pay a substantially below-market interest rate on deposits. If it did so, banks would go to excessive lengths to minimize their liquidity requirements, putting unnecessary strains on money markets.\[^{61}\] However, as discussed in the next section, if it wanted to, the Fed could adjust its regulations and the expectations of examiners to accept alternatives to reserve balances such as reverse repos or collateral prepositioned at the discount window as satisfactory forms of liquidity. This approach would be similar to the one being taken by the Bank of England. Hauser (2023) notes that banks meet their same-day contingent funding needs by pledging collateral to the BoE’s central bank liquidity facilities.

. . . And The High Costs

When the FOMC evaluated its implementation framework choices in 2018, the Committee noted that the costs of the system were “. . . 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.”\[^{62}\] The Committee observed that these costs were proportional to the size of the balance sheet needed to offer excess reserves. For example, at the November 2018 FOMC meeting, Vice Chair Dudley stated “. . . the biggest tradeoff posed by the current system concerns the potentially large size of our balance sheet needed to support it.”\[^{63}\]

One of the main concerns expressed about maintaining a large balance sheet is that it would undercut the Committee’s commitment to reverse its past quantitative easing programs and potentially make future use of QE more difficult. Chair Powell stated at the December 2018 meeting:

Last year, we began the process of normalizing our balance sheet. And the Committee has long emphasized that we would eventually return the balance sheet to a size that’s no larger than it needs to be to conduct policy in our chosen framework. I do see this as a high-profile commitment that we need to honor; and doing so, in my view, would do much to support the legitimacy of future large-scale asset purchases—by showing that QE is not just a one-way street to an ever-larger Federal Reserve balance sheet relative to GDP. . .

I do think, in the end, [better communication] will not be sufficient to convince those who have concerns about QE as implying an ever rising balance sheet as a percentage of GDP. So I do think it’s necessary to do what we can to be seen to have worked hard to

\[^{61}\] Nelson (2016).
get to a smaller balance sheet. It’s not just that we get down to scarce reserves, it’s that we be seen to work hard to do it—that we take the commitment seriously. In effect, to me, that means we “reserve” judgment, in a sense, until we have really made further progress.64

Indeed, in November 2018 several participants stated that if the size proved to be “considerably higher than anticipated” (pp. 3–4), the Committee should revisit their decision. Recall that the staff estimate at the time was that reserve balances of $1 trillion would be needed to operate the excessive-reserves approach. At the December 2018 FOMC meeting, Chair Powell stated:

I think what I would be concerned about is a case in which we find, counterfactually or counter to our expectations, that equilibrium reserve demand is $1.5 trillion or something like that, and then I think I would have buyer’s remorse if we had made a formal decision to go ahead.65

Given these concerns, one reason why the Committee may have decided in January 2019 to officially adopt a floor system is that the staff estimated that the size of the balance sheet in a floor system would not be very different from the size in a corridor system. As noted, staff estimated that $1 trillion in reserve balances would be needed to implement the excessive-reserves approach, but they concluded that $700 billion would be required to implement the necessary-reserves approach—not a large difference. In the subsequent discussion, Patrick Harker, president of Federal Reserve Bank of Philadelphia, stated that “. . . the likely difference in the size of the balance sheet that would result from operating in a corridor . . . and the size of the balance sheet that is no bigger than necessary to ensure a floor system, I think, is unlikely to be significant.”66

The size of the balance sheet necessary to operate a floor system has not only proven to be vastly larger than anticipated, but there is also good reason to think the necessary size will continue to grow over time. The Federal Reserve’s plan is to offer a quantity of reserves sufficiently far above the necessary level so that there can be large negative shocks to the supply (primarily owing to increases in the TGA) without requiring offsetting open-market operations. The Fed estimates the necessary buffer to be about $350 billion.67 But because banks and bank examiners will get used to that higher level of reserve balances (especially because the IORB rate will tend to be rich compared with other money market rates), the necessary quantity of reserves will move up to the original necessary amount plus the buffer.68

As a consequence, the Fed will need to shift reserves up to be $350 billion above the now-higher level of necessary reserves. And so on. The ratchet effect can be seen in the relationship shared by the spread between the fed funds and IORB rates and the level of reserves between

64 Federal Open Market Committee (2018c), pp. 43-44.
65 Federal Open Market Committee (2018c), p. 44.
66 Federal Open Market Committee (2018b)
January 2009 and November 2021 (Exhibit 12). Since November 2021, however, the spread has barely changed even as the level of reserve balances declined, suggesting that not all of the massive increase in reserves caused by QE4 got locked into bank practices and supervisors’ expectations.

![Exhibit 12: Spread Between the Fed Funds Rate and the IORB Rate](image)

Another serious cost of an excessive-reserves system is that it degrades interbank markets, where banks turn to meet liquidity needs in the private sector without government involvement. When the commercial banking system is massively overstuffed with reserve balances, a federal funds market is no longer needed, and the market deteriorates. Before the GFC, large banks borrowed and lent to each other daily on an unsecured overnight basis in the fed funds market. As discussed by Selgin (2018), because an institution generally had to be of unquestioned quality to borrow in the market, and losing access would be devastating for the bank, access proved to be a continuing source of market discipline.

As we noted, the fed funds market currently is almost entirely loans from Federal Home Loan Banks to the U.S. branches of foreign banks, with the fed funds rate consistently trading a bit below the IORB rate. FHLBs have accounts at Federal Reserve Banks but do not earn interest. The FHLB banks are allowed to count the fed funds loans as a source of liquidity by their supervisor, the Federal Housing Authority. The branches borrow from the FHLBs at a rate a bit below IORB rate and simply lend the funds to the Fed as reserve balances, turning a profit. The borrowers tend to be branches of FBOs, because they do not have insured deposits and therefore no deposit insurance premiums. Such premiums are based on total liabilities.
(including fed funds borrowings), not just insured deposits, so if a bank participated in the arbitrage transactions, its deposit insurance premiums would go up.

The subsidized reserve balances have also degraded sources of liquidity for banks from the Federal Reserve. Although borrowing from the discount window has been stigmatized since the 1920s, use has declined sharply under the excessive-reserve framework, which has likely fomented the stigma even more. Bank treasurers report that they are extremely reluctant to borrow, and if they did, their examiners would expect to be immediately informed.69

The abundance of reserves has also resulted in there being a stigma associated with running a collateralized daylight overdraft. Before the GFC, such overdrafts drew no stigma. Indeed, in 2011, the Fed revised its daylight overdraft policy to make collateralized overdrafts free and essentially unlimited. The objective was to reduce any incentive for banks to hold back payments during the day, removing an important source of systemic risk. The new stigma associated with daylight overdrafts has reportedly been boosted by examiner attitudes. Banks are not allowed to anticipate making a daylight overdraft in their resolution plans, so making an overdraft calls into question the credibility of those plans.

Using the framework has also required a significant increase in the Fed’s involvement in the financial system. Under the necessary-reserves system, banks kept the minimum amount necessary on reserve at the Fed, and reserves were 0.1 percent of bank assets.70 As discussed, the Fed conducted only small, relatively unimportant repo operations with the primary dealers, and a bank took out an occasional discount-window loan. As of November 2023, reserves are $33 trillion and represent 14 percent of banks assets, and ON RRPs are $1 trillion and make up 17 percent of money fund assets.71 Treasury is investing $375 billion in the Fed through the TGA (and planning on raising that amount substantially higher72), and foreign institutions are lending over $300 billion to the Fed through reverse repurchase agreements.

Somewhat more subtly, the use of a floor system has contributed to a view at the Fed that getting bigger essentially carries no costs. That view has increased the Fed’s willingness to engage in asset-purchase programs and conduct emergency lending. Lorie Logan, president of the Dallas Fed, stated in a 2023 speech: “I don’t see large costs of supplying the quantity of reserves needed to establish a liability-driven floor.”73 In 2008, Bill Dudley, then president of the New York Fed, noted that a shortcoming of a corridor system compared with a floor system is precisely that it slowed the Fed down:

---

71 At the beginning of November 2023, reserve balances equaled $3.3 trillion and commercial bank assets equaled $23.0 trillion. See https://www.federalreserve.gov/releases/h41/20231102/ and https://www.federalreserve.gov/releases/h8/20231103/ For money fund assets, see https://www.ici.org/research/stats/mmf.
73 Logan (2023).
Also, I think a corridor regime has one important negative aspect that maybe could be highlighted a little bit more. In 2007 and 2008, we were reluctant to enact large programs, and we didn’t enact programs that were open-ended, because we were worried about how we would actually offset, by corresponding draining operations, the reserves added through such programs.74

Adopting a floor system has also eliminated a safeguard on the Fed’s balance sheet being used by Congress as a source of financing.75 If asked to buy this or that security, previously, the Fed could demur by correctly observing that it would lose control of monetary policy if its assets exceeded currency by more than a small amount. At the November 2018 FOMC meetings, Loretta Mester, president of the Cleveland Fed, stated “The lack of an operating constraint on the size of our balance sheet might also generate requests that the Federal Reserve aid specific industries or use the balance sheet to fund government initiatives, as occurred during and since the crisis.” Similarly, at the same meeting, Randal Quarles, vice chair for supervision, stated:

Having the FOMC control such a large stock of assets presents what the lawyers in the room will recall from your first-year torts class is called an “attractive nuisance.” And for the nonlawyers in the room, an attractive nuisance is an object that a property owner allows to remain on his land when it is obvious both that the object will be dangerous if misused and that misusing it will be irresistibly appealing to passers-by of impulsive and immature judgment, such as children and congressmen. 76

Subsequent experience has shown that this is not just an abstract concern. The financing plank of the Green New Deal is essentially this:

As the checks go out, the government’s bank—the Federal Reserve—clears the payments by crediting the seller’s bank account with digital dollars. In other words, Congress can pass any budget it chooses, and our government already pays for everything by creating new money.77

Similarly, Saule Omarova, a 2021 nominee to head the OCC, proposed that the Fed offer everyone accounts. It could then put money into the accounts of businesses if they retained their employees and spent money on “real” goods and services as well as the accounts of underprivileged people, expenditures the Fed would finance by driving its equity negative.78

The ECASH Act proposed in 2023 would direct the Treasury to create a digital currency, with costs covered by running an overdraft in a specially created account at the New York Fed.79 Somewhat more subtly, the CARES Act of 2020, after encouraging the secretary of the Treasury to implement a program to aid medium-sized businesses, states: “Nothing in this subparagraph

74 Federal Open Market Committee (2016).
75 See Plosser (2022) and Selgin (2020).
76 Federal Open Market Committee (2018b).
78 Omarova (2021).
79 ECASH Act (2022).
shall limit the discretion of the Board of Governors of the Federal Reserve System to establish a Main Street Lending Program or other similar program or facility that supports lending to small and mid-sized businesses. . .”\(^8\)

Relatedly, the floor system has also opened the door to the Fed monetizing the debt. It is unclear what “monetizing the debt” actually means. But in 2011, Chairman Bernanke was asked by the House Budget Committee if the Fed was monetizing the debt with its large-scale asset purchases. He responded:

No, sir. Monetization would involve a permanent increase in the money supply to basically pay the government’s bills through money creation. What we are doing here is a temporary measure which will be reversed so that at the end of this process, the money supply will be normalized, the amount of the Fed’s balance sheet will be normalized, and there will be no permanent increase, either in money outstanding, in the Fed’s balance sheet, or in inflation.\(^8\)

Recall that at that time, the Fed’s normalization principles involved selling the securities acquired during QE to return to a necessary-reserves framework. But the Fed has instead decided to maintain a giant balance sheet, much larger as a percentage of GDP than when Bernanke said there would be no permanent increase. By Bernanke’s definition, therefore, the Fed is monetizing the debt.

More broadly, although the Fed will never deliberately set out to monetize the debt, with the balance sheet unbounded, it is free to respond to developments in a manner that does so. For example, Vice Chair Quarles explained in October 2020, the Fed may need to continue massive securities purchases, because the Treasury was issuing more than financial markets could handle on their own.\(^8\)

**How the Fed Could Return to a Necessary-Reserves Implementation System**

The Federal Reserve can return to a corridor system where the discount rate was materially above and the IORB rate materially below the fed funds rate. The Fed is currently reducing the size of its balance sheet by allowing maturing Treasury securities and agency MBS to mature without reinvestment. To get up off the floor, the Fed needs to continue that process until money market rates move somewhat above the IORB rate. At that point, as in 2019, banks will begin to replace reserve balances with other investments (such as reverse repurchase agreements) as a source of liquidity. JPMorgan Chase made such a switch in mid-2019. JPMorgan CEO Jamie Dimon commented then on the shift:

. . . we have a checking account at the Fed with a certain amount of cash in it. Last year we had more cash than we needed for regulatory requirements. So when repo rates

\(^8\) CARES Act (2020).
\(^8\) Derby (2020).
went up, we went from the checking account, which was paying IOER into repo. Obviously makes sense, you make more money.\textsuperscript{83}

Over time, examiners and bank management will get comfortable with lower levels of reserve balances, and the process will be able to continue.

This is the approach the Bank of England is currently using. It plans to shrink its balance sheet until borrowing at its standing repo facility picks up. At that point, money market rates will probably be a bit above Bank rate, encouraging reductions in demand for reserves.\textsuperscript{84}

Moreover, the reduction need not result in another episode of money market volatility, such as in September 2019. There was nothing remarkable about the circumstances that led to the turmoil that September. Treasury securities settlement and corporate tax day coincide every quarter. What was unprecedented was the combination of (1) reserves beginning to be scarce and (2) the Fed not intervening to prevent a sharp decline in reserves. In the years after the GFC, reserves were so excessive that the tax day/settlement day declines in reserves could be absorbed without turmoil. Before the GFC, the tax receipts would have flowed back into the banking system, and the Fed would have offset any residual reduction in reserve balances with an open market operation.

Indeed, the turmoil in September 2019 was well anticipated. The Bank Policy Institute wrote \textit{two weeks before the turmoil}:

> Because Treasury tax receipts will be building up and Treasury securities issuance will be heavy, banks’ reserve balances at the Fed are going to decline rapidly and significantly to $1¼ trillion from their current level of $1½ trillion. As a result, money market rates will become volatile as banks scramble for reserves.\textsuperscript{85}

Similarly, Lou Crandall (2019) at Wrightson ICAP described what was going to happen in considerable detail \textit{a week before the turmoil}:

> The TGA will spike on Monday, September 16 due to quarterly tax payments and the settlement of the Treasury’s mid-month coupon auctions. . . . In the near term, the effects of the run-up in the TGA are likely to put upward pressure on overnight rates in both the secured and unsecured market. That will probably have less to do with the decline in reserve balances per se than the reasons why reserve balances are falling. The money flowing into the Treasury’s account will flow out of banks and the repo market. (Some of the outflow will come from money fund balances, which will reduce the amount of funding available to the repo market.) That funding drain . . . is likely to push short-term rates higher.\textsuperscript{86}

\textsuperscript{84} Hauser (2022).
\textsuperscript{85} Nelson and Waxman (2019).
\textsuperscript{86} Crandall (2019).
Nevertheless, writing two years after the event, staff at the New York Fed emphasized how unexpected the 2019 turmoil was. Indeed, the first line of their report is “U.S. money markets unexpectedly experienced severe upward rate pressure on September 16 and 17, 2019” [emphasis added]. The report goes on to say:

This episode attracted considerable attention among market practitioners, academics, and policymakers, in part because it was unexpected. On September 16, two transitory shocks hit money markets: the quarterly corporate tax payment and the settlement of the mid-month Treasury coupon auction. Money market observers were well aware that the occurrence of both events on the same day would lead to a drop in the level of reserves, a reduction in cash available to be lent in the repo market, and an increase in the demand for cash in that market.

On their own, however, these two shocks do not provide a compelling explanation for what happened, both because they were well anticipated and because shocks of a similar type and magnitude had happened before without significant rate pressures.\(^{87}\) [emphasis added]

It was certainly no secret that there were increasing signs that reserve balances were becoming somewhat scarce in September 2019. The federal funds rate had risen above the IORB rate in April 2019. Because reserves were becoming scarcer and there was a well-forecasted sharp drop in reserves on Sept. 16, it was the responsibility of the New York Fed (working with Board staff) to conduct an open-market operation to offset the decline. Future episodes of similar volatility can be avoided simply by doing what the Fed had always done: adding reserves using open market operations when autonomous factors are projected to reduce them.

As reserves decline further and the fed funds rate rises notably above the IORB rate (perhaps 50 to 100 basis points above), the fed funds market will recover. Banks will once again have a strong incentive to keep their reserve balances to the minimum necessary amount, and so be eager to lend excess at the end of the day. Anbil and Carlson (2019) describe how the funds market reemerged in the 1950s as the Fed reduced reserve balances from a super-abundant level.\(^{88}\) They conclude, “This historical experience may have implications for current monetary policy because it suggests that as the Federal Reserve reduces the supply of reserves through tightening monetary policy, and as the incentives to participate in the market return, interbank trading in the federal funds market is likely to eventually revive.” (p. 8)

The process of returning to a corridor system would be helped if the Fed returned to announcing a point target for the federal funds rate, rather than a target range. Doing so would strengthen the influence of the Fed’s announcement on fed funds market participants’ coordination on a new level for the fed funds rate.

---

\(^{87}\) Afonso and colleagues (2021), pp. 1–2.

\(^{88}\) Anbil and Carlson (2019).
If the Fed wished, at some point before money market rates rose above the IORB rate, it could implement a voluntary reserve requirement regime, as we discussed. Doing so would give the Fed good interest rate control using a corridor system while still allowing each bank to choose to hold the level of reserves it wants when reserves are compensated at a market rate. As Baughman and Carapella (2019) note, such a regime would reinvigorate the federal funds market. The availability of ex-post liquidity insurance could reduce banks’ demand for reserves and so allow the Fed to cut the size of its balance sheet over time. If the Fed undersupplied reserves slightly so that discount-window borrowing occurred regularly, banks would have an incentive to cut their reserve balances over time, reducing the size of the Fed. Moreover, the regular borrowing would itself help ease stigma.

One challenge to returning to a floor system is lessening the volatility of the TGA. Presumably the Fed could again ask the Treasury to limit the size (or at least the volatility) of its cash balance, investing its excess cash at banks or in the repo market. Other central banks are taking similar actions. In February 2023, the ECB reduced the interest it pays on Euro-area government deposits by 20 basis points to encourage a gradual decline in those deposits. The ECB stated that the change reflected “the desire to encourage market intermediation. . . .” At the end of February 2023, the Bank of England had only 4.3 billion pounds on deposit from HM Government accounts. Of the U.K. government’s financial assets, only 1.5 percent were in deposits at the BoE as of March 31, 2023.

Another possibility would have the Fed segregate the TGA and conduct daily repo operations to match the size of the account expected at the end of the day. Recalling that when the Fed conducts a “repo” it is really conducting a reverse repo, the Treasury would be investing funds with the Fed, and the Fed would in turn be investing the funds in the repo market. Although the repo operations would have to be large, the TGA’s volatility has to show up somewhere. The idea that the supply of reserves can simply move back and forth smoothly along the flat part of the demand curve without consequences has proven incorrect.

In the end, though, the Fed can get no smaller than the size needed to create the reserve balances banks are required (by regulation and their examiners) to hold. That level is likely to be much larger than the de minimis levels banks deposited at the Fed before the GFC.

That level can be reduced by encouraging banks to be prepared to meet their contingency funding needs with collateral prepositioned at the discount window as a substitute for reserve balances. For example, a report released by the Group of Thirty in January 2024 suggests that banks be required to have borrowing capacity at the discount window when combined with reserve balances equal to uninsured deposits and some other types of wholesale funding, and the Acting Comptroller of the Currency Michael Hsu has proposed that such capacity be at least

---

89 Hauser (2019).
90 ECB (2023).
92 United Kingdom Debt Management Office (2023).
equal to the bank’s cash needs under five days of severe liquidity stress. In a podcast in January 2024, Borio described the advantage of returning to a corridor system with an active federal funds market and looking to discount window borrowing capacity as a bank’s source of contingent liquidity. In that case, banks first meet their liquidity needs in the interbank market, then prepare for their liquidity needs under stress with collateral at the discount window, as opposed to the current system in which banks meet their liquidity needs with deposits at the Fed:

Would you like to have a system in which the central bank is a backstop, or would you like to have a system in which the central bank is the mass market maker of first resort, so last resort versus first resort?... I think that having a system in which the central bank is a backstop, and a system in which the first line of defense against demands on liquidity is an interbank market, that to me sounds [like], on balance, a better system.

**Conclusion**

The Federal Reserve’s shift to an excessive-reserves framework, initiated in response to the 2008 financial crisis and solidified in January 2019, has proven to be a mistake. The original intention was to return to the necessary-reserves approach. But a series of developments, including multiple waves of quantitative easing, led to the entrenchment of the excessive-reserves system.

Contrary to its purported benefits, the excessive-reserves framework has not simplified monetary policy. Although the federal funds rate has remained stable, the federal funds market has essentially disappeared. Implementing the system has required that the Fed operate with a much larger set of counterparties and maintain a much larger footprint in the financial system.

Advocates argued that the system was superior because it aligned with the Friedman rule, suggesting that an abundance of reserves is socially optimal. However, the Friedman rule can be satisfied in a corridor system using the originally recommended approach from April 2008: a voluntary reserve regime paying a market rate on self-required reserves.

The costs associated with the excessive-reserves system are significant. The size of the Federal Reserve's balance sheet has expanded dramatically, and the system lacks a natural limit. This contributes to the misconception that an increased size comes at no cost. Moreover, there is a ratchet in the demand for reserves—the more the Fed supplies, the more banks demand. They have adjusted their operations and balance sheets to accommodate the abundant and inexpensive reserve balances. Consequently, the minimum level of reserves the Fed judges necessary to conduct policy effectively in a floor system has risen 100-fold over the past 15 years.

---

93 Group of Thirty (2024) and Hsu (2024).
94 Borio (2024).
The Fed should return to a necessary-reserves system using a corridor framework. To achieve this shift, the Federal Reserve should reduce the size of its balance sheet, prompting money market rates to rise and making it more expensive for banks to lend excess reserves. The ratchet effect would gradually diminish, allowing the necessary amount of reserves to shrink over time.

And once the level of reserve balances approaches the current minimum level that banks demand, the Fed needs to once again control volatility in reserve balances using open market operations. The Fed’s failure to do so led to the significant bout of repo market volatility in September 2019.

As part of the transition, the Fed could implement a voluntary-reserve-requirements framework, allowing banks to set their reserve levels over a maintenance period and compensating them at the FOMC’s target federal funds rate. Excess reserves would be compensated at a lower rate, and reserve shortfalls penalized, replicating the cost of borrowing from the discount window.

By guiding reserves to a point where discount-window borrowing picks up and the fed funds rate averages slightly above the interest rate the Fed pays on voluntary reserve targets, the Fed could encourage banks to economize on their use of reserves and reduce the social costs of operating with a large balance sheet. Although reserve balances cannot get lower than the level examiners require banks to hold for liquidity risk management purposes, that level can be reduced substantially by encouraging banks to meet their short-term cash needs under stress in large part with capacity to borrow from the discount window. The return to a necessary-reserves system, as outlined in this paper, offers a path toward good control of interest rates, along with a much smaller Fed with a greatly reduced footprint in the financial system.
References


https://doi.org/10.17016/FEDS.2020.022


https://www.newyorkfed.org/newsevents/speeches/2017/log170518


https://bpi.com/this-is-not-what-transparency-looks-like/


