Payment Needs and the Size of the Federal Reserve's Balance Sheet*

Ricardo Reis

LSE

March 2026

The Federal Reserve's (Fed) balance sheet has changed dramatically since 2007. Its size grew from 6.1% of GDP back then to 21.6% of GDP by the end of 2025. The main item on the liabilities—deposits of banks at the Fed, commonly called reserve balances—was $2.9 trillion at the end of 2025 paying a total interest bill of $12.1bn, down from $68bn in 2024, but up from $0 before 2007, when these reserves paid no interest.

Now, in 2026, the incoming chairman of the Fed, Kevin Warsh, has declared his desire to shrink the size of the balance sheet. How far can it go? Will doing so cause volatility in financial markets? Are there policy changes that can ease possible negative effects of this shrinkage?

The ABCDs of the demand for reserves to make payments

The primary use of reserves is to make payments. Consider the narrowest possible bank in order to focus solely on its payments activities. This bank holds as assets Treasury bills (T-bills) b and reserves v, which pay interest rates i^b and i^v, respectively. While T-bills are liquid, safe, and desirable on all accounts, they cannot be used directly to make payments; they must be sold first. Reserves share all the desirable features of T-bills, but they *can* be used for payments. This service benefit allows reserves to pay a weakly lower interest rate than T-bills: $i^v \leq i^b$.

A bank allocates investments between T-bills and reserves knowing that payment

*Contact: r.a.reis@lse.ac.uk. These comments were delivered at the Spring conference of the *Brookings Papers on Economic Activity* as a discussion of Duffie (2026). I am grateful to Saleem Bahaj for discussions. This work was supported by the UK Research and Innovation grant number EP/Eo25039/1.

demands will arrive at any time. When it makes a payment, the bank's reserves v fall into a deficit, while the reserves of the receiving bank rise into a surplus. The bank's optimal reserve holdings depend on four factors: A, B, C, and D.

A is a regulatory minimum—perhaps hard, usually soft—on reserve holdings. If the reserves of a deficit bank fall too close to this minimum, it suffers penalties and may even lose its banking license.

B comes from banks with a surplus of reserves having an unmet demand for reserves from beforehand. They value the additional inflow of reserves and prefer to hold on to them.

C is the bargaining process between surplus and deficit banks, which determines the terms of trade between them, or the money market (or interbank) rate.

D comes from the option that deficit banks have to borrow from the discount window, at a cost that may include not only the posted rate of interest but also a shadow penalty reflecting how other banks and supervisors adversely view such borrowing.

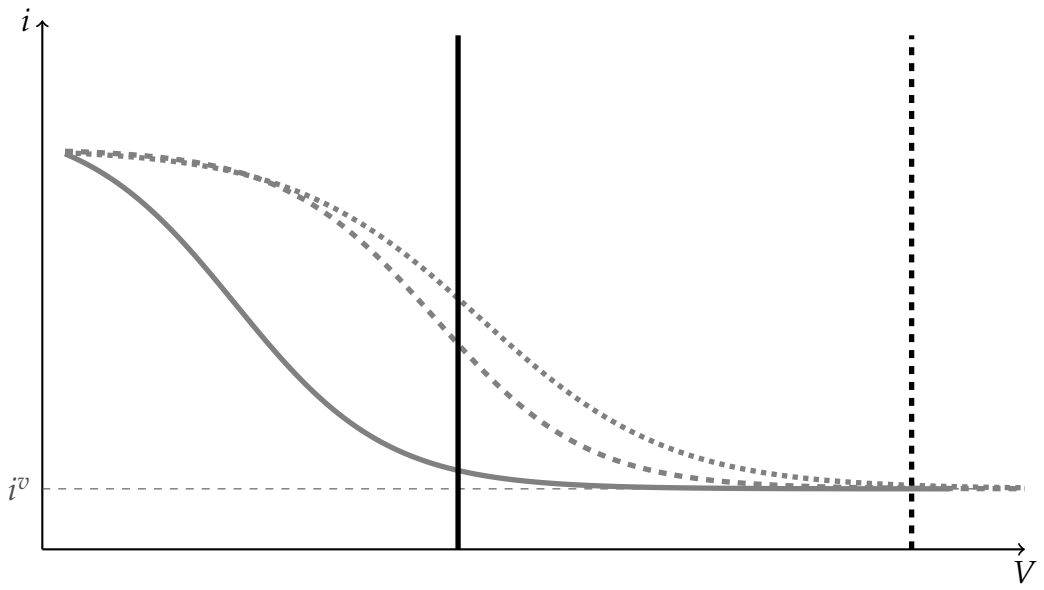
These four factors generate a demand for reserves v that depends on the interest rates i^b and i^v, as well as on the factors A, B, C, and D. This is usually called a Poole model of liquidity after the seminal work of Poole (1968); see Bahaj and Reis (2026) for a modern derivation.

Graphically, this maps a downward-sloping demand curve by the bank for reserves against the money market rates, bounded below by the interest on reserves, and bounded above by the cost of using the discount window. Adding up across many banks to get a total demand V that is the sum of the individual v's, each with A, B, C, and D's working with different strengths, produces the smooth aggregate demand curve plotted in figure 1 in solid gray.

The problem and the solution

Duffie (2026) carefully explains that, since 2007, this curve shifted outwards (see also Anderson et al. (2026)). First, liquidity regulations increased the soft minimum of reserves that banks have to hold as "own liquidity". Second, surplus banks became less willing to lend, partly because the interest rate gap $i^b - i^v$ was smaller as a result of reserves paying interest. Third, frictions in interbank matching have worsened and tilted the bargaining

Figure 1 The money markets with an inelastic supply of reserves



terms towards the surplus banks, partly because of FDIC charges that are incurred by banks participating in this market. Fourth, and finally, the penalty associated with borrowing from the central bank has increased, as the larger banks have been discouraged by supervisors from intraday borrowing from the Fed. Respectively, A, B, C, and D all shifted. Moreover, since each of these factors depends on different regulations binding and different supervisors exerting their suasion, then the curve also became more volatile. Figure 1 shows this new demand curve in both usual and elevated times, as dashed and dotted gray curves, respectively.

The Fed chooses to supply reserves inelastically. It sets a quantity of reserves it wants to issue and passes them on to banks in exchange for their Treasuries. Combined with the new demand curve that has shifted right and become more volatile, this creates a problem.

With its old, small balance sheet, represented in figure 1 by the black vertical supply curve, money market rates would be well above the interest paid on reserves. The Fed could shift supply to the right in order to hit the desired money market rate that the FOMC members decide is adequate to achieve its target for inflation. But calibrating where to place the vertical line to intersect demand at just the right point was already quite hard pre 2007, and is much harder with a demand curve that is so far to the right. Moreover, the higher volatility of demand makes it inevitable that often the intersection happens at dramatically higher rates. Volatility in money markets becomes the norm, and with it the

dangers of a financial crisis because some banks are unable to make their payments or suffer large losses in the process of securing the reserves they need.

There is a further reason why this situation is undesirable. The Friedman rule persuasively argues that, since a central bank can produce reserves at close-to-zero marginal cost, then the opportunity cost of holding reserves to make payments, which is captured by the difference $i^b - i^v$, should be close to zero as well. Therefore, the intersection point should be far to the right. If the demand curve has a kink point, becoming horizontal to the right of that point, then that kink is the Friedman rule optimum.

The figure makes the solution to the problem obvious: shift the supply curve far to the right to the black dashed line. Have a large balance sheet providing lots of reserves to the banks. The market rate would be close to the administrative interest on reserves at almost all times, making the conduct of monetary policy easy. Volatility would be minimal. The market for payments would be saturated, as demand would be satiated. I characterized this solution in Reis (2016), and it has been given many names: structural liquidity surplus, abundant reserves, floor system, among others.

For most of the past 17 years, this is the world that we lived in, largely as an unintended consequence of quantitative easing. Looking ahead, some central banks—notably the Norges Bank and the Bank of Canada—have chosen to remain in this regime for the future.

The new problem and a new solution

The incoming president of the Federal Reserve has made a case for reducing the size of the Fed's balance sheet (Warsh, 2025). I will not dispute this goal or discuss the arguments in favor or against it. At the same time, I will also take as given that the Fed will keep to the two other goals: not deviate too much from the Friedman rule, and avoid volatility of money market rates that could trigger a financial crisis. Sometimes, Fed officials refer to this as having "ample reserves".

How can it achieve this? Figure 1 again makes the answer crystal clear: shift the demand curve back to the left. There really is no other way. Policies to reduce the balance sheet are policies that reduce the demand for reserves.

The Fed perfectly controls the supply curve, so shifting it to the left is so easy as to barely merit discussion. When discussing how to reduce the balance sheet, the discussion must be all about how to shift the demand curve. That is what may be hard, but also what is essential.

Duffie (2026) proposes three demand-shifting policies (see also Logan and Schulhofer-

Wohl (2026)). The first is to reform or scale back some of the regulations that have, explicitly or implicitly, increased the soft minimum level of reserves at which a deficit bank runs into trouble with its supervisor. Shift A partly back. The second is to introduce a tiering remuneration of reserves so that surplus banks are eager to lend out their reserves. Shift B partly back. The third is to introduce liquidity savings mechanisms that reduce the volume of trading and matching in the interbank market. Shift C partly back.

A, B, and C may have shifted right after the great financial crisis for valid regulatory and macro-prudential reasons. But, given the new goal of shrinking the size of the balance sheet, these have to give in—by a little or a lot—to bring A, B, and C back down.

Shifting the demand is not enough to prevent volatility

However, relaxing or modifying these regulations is unlikely to significantly reduce the volatility of demand. The rules, and their complex interaction with each other, will still be there. Moreover, moving the demand curve will inevitably not be a smooth process that can be perfectly foreseen by the Fed. In the transition to a slimmer balance sheet with less demand for reserves, banks will have to revert habits and business models from the last decade. Even if only temporarily, the demand will be moving around and, if the supply curve is shifting to the left, money market volatility will result. A piece of the policy answer is missing.

One extra policy to fill that hole is to adjust the vertical supply in response to every shift in demand. This would be implemented, as usual, with shifts in the inelastic supply executed by open market operations: buying and selling bonds and in exchange paying or receiving reserves. Insofar as these shocks to demand are temporary, these may be described as temporary open market operations.

Note that describing this as a "supply driven system" is accurate but also misleading. Yes, the Fed would be changing the supply of reserves. But, in every instance, it would be doing so in response to a shock to demand. Demand is what, in a causal sense, would be driving the level of reserves and the size of the Fed's balance sheet.

No matter what one calls it, this could reduce the volatility. But, surely, shifting the supply curve around to keep track of shifts in demand will come with errors. As good as the Fed staff is with its temporary buying and selling of bonds, identifying the demand shocks in real time, and responding to them right away is not an easy task. It is likely that significant volatility will remain.

The elephant in the room

However, there is an elephant in the room that the figure again makes transparent. Why should you have a vertical supply curve in the first place at all?

In fact, the Federal Reserve Act of 1913 that created the Fed explicitly instructed it "To furnish an elastic currency." In our digital age, reserves have taken the role that physical currency had back then, but the principle remains. As the legislators more than a hundred years ago understood, having a vertical supply curve inevitably leads to financial instability.

An elastic supply of reserves is a supply of reserves that is vertical at some level, call it V^*, and horizontal to the right of that at a rate i^*. Figure 2 displays it in black.

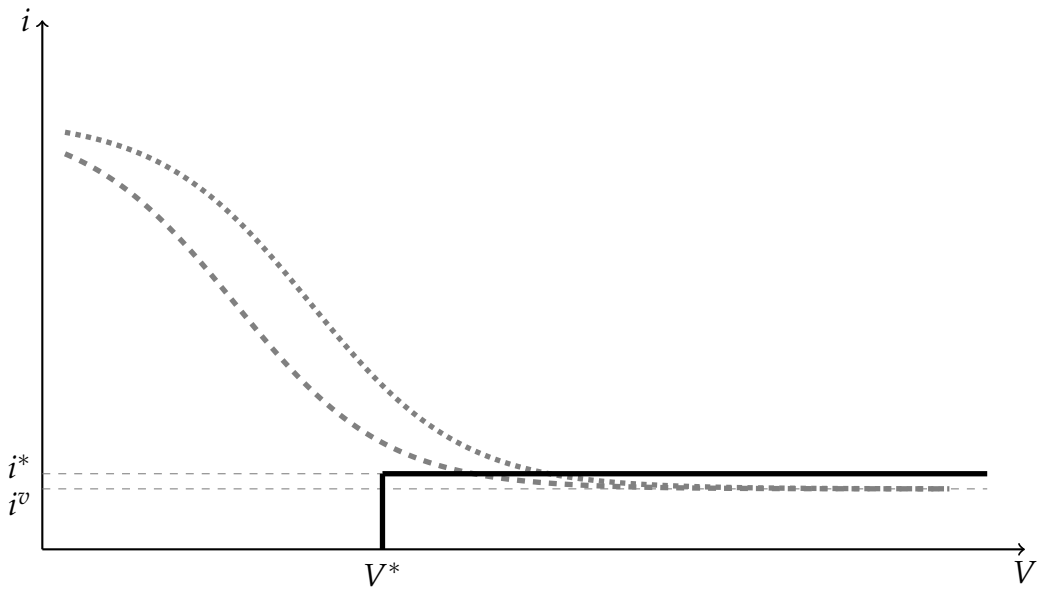
How is it created in practice? The Fed would have a steady amount of deposits by the banks V^* achieved through open market operations. Then, it would run a standing repo facility with full allotment at a fixed rate i^*. What this means is that a bank could come to the Fed and exchange its bills b for reserves v when it anticipates having a liquidity deficit. They could have this extra demand in whatever size (full allotment) at a fixed announced rate i^*.

This repo facility would be open infrequently, perhaps daily or weekly, with the policymaker choosing the desired frequency given how much volatility of money market rates is deemed to be desirable. When the repo facility is open, as figure 2 shows, the money market rates would all cluster at the desired policy rate i^*. If it is closed, the money market rates would fall towards the interest on reserves i^v when demand for reserves is weak, and rise to the discount window penalty rate when demand is strong. This arrangement is sometimes called a corridor system.

By choosing the difference $i^* - i^v$ to be a small number (say 10 or 25 basis points), then policymakers can speak of these two rates—on reserves and on repos—interchangeably as being the policy rate. Then, by picking the level of V^*, they can control the volatility of money market rates. This is because the further the vertical segment of the supply curve is to the right, the rarer will be the times when the repo balance is positive, and interest rates deviate upwards. Whether the choice of V^* makes it be to the right or to the left of the normal (or average) level of reserves at i^* is sometimes described as having a structural surplus or deficit of reserves, respectively.

It is for the policymaker to decide these parameters of the system. The essential feature is that reserves are supplied elastically. Demand still drives what the equilibrium level of reserves is. So, it is still the policies that shift A, B, C, and D that determine the size of the

Figure 2 The money markets with an elastic supply of reserves



balance sheet. Having an elastic supply allows the shrinkage of the Fed's balance sheet to happen without financial crises in money markets.[1]

The virtues of an elastic supply of money

The volatility of the demand curve is partly driven by daily shifts in payment needs in the economy and by seasonal and business cycles, and partly by the inevitable errors in estimating its present location. Aside from restricting the resulting volatility of money market rates, an elastic supply curve has the virtue that it keeps the average size of the balance sheet as small as possible. If the policymaker chooses an $i^* - i^v$ sized deviation from the Friedman rule, then the intersection of supply and demand will be at the lowest possible level of reserves that just achieves it. Trying to target this level with an inelastic supply would often lead to being to the right of this minimum, with a lower $i^* - i^v$ and a higher V. An elastic supply curve guarantees the smallest possible size.

In terms of the composition of the balance sheet of the banks, there is no difference between elastic or inelastic supply. When the demand for reserves increased, open market operations lowered T-bills and raised reserves. The repo operations do the same, and

[1] In fact, back to the figure, frequent open market operations that keep on shifting the supply curve around the flat segment of the demand curve are tracing out a horizontal line of outcomes, albeit imperfectly and noisily relative to what a horizontal supply curve can achieve. See Afonso et al. (2025) for a formal analysis.

likewise when the demand fell. Therefore, in terms of what should matter for regulation, there should be no difference.

The objection to having a standing repo as an elastic supply function almost always collapses to one word: stigma. Banks would be reluctant to exchange their T-bills for reserves even when they needed it. Note that they are happy to do these exchanges to buy and sell T-bills, but they would be averse to doing the exact same exchange, leading to the same balance sheet, if it was done through a repo.

Why? What is this stigma? Why would ruthlessly profit-maximizing banks leave millions on the table by borrowing reserves at market rates above what they would pay for a repo at the central bank's standing facility?

Those who argue for stigma in using the repo facility make an analogy to how banks avoid borrowing from the discount window. But, borrowing from the discount window is done infrequently; using the repo facility would be common. The terms of the discount window include a penalty; the repo facility uses the policy rate, and the rate paid on the reserves held via repos would be quite close to the deposit rate paid on reserves held outright. The few banks using the discount window are identified by the other banks as being in some trouble; most banks would use the repo facility.

The Fed created such a facility in December of 2025—just a few months ago—that is open every morning, and with an interest rate 10bp above the interest it pays on outright reserves (for two previous years, it had experimented with a fixed allotment facility, which is an inelastic supply, so not the same). Banks have not used it as much as one would expect, sometimes borrowing reserves at higher rates from other banks than what they would pay on their repo balances. It is still too early to declare that banks are irrational and leave millions on the sidewalk. Rather, looking at the behavior of supervisors, and at what bankers say, strongly suggests that the reluctance to use the repo is a symptom of a problem: supervisory failure.

Supervisors, when applying the Comprehensive Liquidity Analysis and Review (CLAR), show a preference for banks to meet liquidity requirements by holding reserves rather than Treasuries. But, if there is a repo facility that lets banks convert T-bills into reserves on demand without limits at a fixed rate, there is no reason to do so.

The Fed does not centrally clear its repo operations. Because of that, when a bank uses the repo facility, this consumes a small amount of capital. There is no justification for this.

According to bankers, supervisors seem to treat usage of the repo facility in the same way that they treat borrowing from the discount window. But they are not the same, the

former being for regular access to reserves, while the latter is for emergency access at a penalty rate.

Stigma is doublespeak for something else: supervisory failure.

The cross-country evidence points in this direction as well. Banks are regulated as much, or more, in the other major currency unions. In almost all of them, there is a repo facility that banks use most of the time. Accordingly, other central banks have opted for supplying reserves elastically (Bailey, 2024, Schnabel, 2025).

Stigma is neither mysterious irrationality by bankers, nor an inevitable consequence of banking regulation; it is a result of how US supervisors implement the rules (see also Ennis, Jarque and Malek (2022)).

Conclusion

After the global financial crisis, banks' need for reserves at the central bank has increased and become more volatile. This demand drives the minimum size of the central bank sheet that is consistent with efficient provision of liquidity (the Friedman rule) and with avoiding liquidity crises in financial markets. As the Fed is near that minimum today, any further reduction in the size of its balance sheet requires policy changes on the side of the demand for reserves, not supply.

However, the shift in demand takes place with humps and bumps. Demand fluctuates across days, weeks, and months with payment needs from firms and households. Economists also struggle to estimate exactly where that demand is. Preventing volatile money market rates requires having a standing repo facility where banks regularly convert T-bills into reserves on demand. This is a necessary condition for the transition to a smaller balance sheet to happen without risking a financial crisis. It is likely necessary even if the balance sheet stays at its current size, since volatility of demand is permanently higher.

The Fed has for a few months had the Standing Repo Operations. These operations have to be perfected to work more smoothly and especially to be included in central clearing. Most urgent is a reform of banking supervision, which currently undermines the efficient functioning of the economy's payment system by penalizing banks for using the Fed's facilities.

References

Afonso, Gara, Gabriele La Spada, Thomas M. Mertens, and John C. Williams. 2025. "The Optimal Supply of Central Bank Reserves Under Uncertainty." *Federal Reserve Bank of New York Staff Reports no. 1077*.

Anderson, Alyssa G., Alessandro Barbarino, Anthony M. Diercks, and Stephen Miran. 2026. "A User's Guide to Reducing the Federal Reserve's Balance Sheet." *Finance and Economics Discussion Series 2026-019*.

Bahaj, Saleem, and Ricardo Reis. 2026. "The Anatomy of a Peg: Lessons from China's Parallel Currencies." *CEPR Discussion Paper 18749*.

Bailey, Andrew. 2024. "The Importance of Central Bank Reserves." *Goodhart lecture at the London School of Economics*.

Duffie, Darrell. 2026. "The Payment System Puts a Floor on the Fed's Balance Sheet." *Brookings Papers on Economic Activity*.

Ennis, Huberto M., Arantxa Jarque, and Hoossam Malek. 2022. "Bank Resolution and the Fed's New Standing Repo Facility." *FRB Richmond Economic Brief 22-06*.

Logan, Lorie, and Sam Schulhofer-Wohl. 2026. "Options for reducing the size of the Fed's balance sheet." *FRB Dallas In Depth*.

Poole, William. 1968. "Commercial Bank Reserve Management in a Stochastic Model: Implications for Monetary Policy." *Journal of Finance*, 23(5): 769–791.

Reis, Ricardo. 2016. "Funding Quantitative Easing to Target Inflation." In *Designing Resilient Monetary Policy Frameworks for the Future*. Jackson Hole Symposium:Federal Reserve Bank of Kansas City.

Schnabel, Isabel. 2025. "Towards a New Eurosystem Balance Sheet." *Speech at the ECB Conference on Money Markets*.

Warsh, Kevin. 2025. "Commanding Heights: Central Banks at a Crossroads." *Annual G30 lecture at the IMF Spring Meetings*.