

THE FED'S OPERATING FRAMEWORK: HOW DOES IT WORK AND HOW WILL IT CHANGE?

Stephen D. Williamson

In September 2008, during the global financial crisis, the Federal Reserve commenced an unprecedented program of asset purchases. At that time the Fed's total assets were about \$925 billion, and when the balance sheet expansion ceased in October 2014, total assets stood at about \$4.5 trillion. Before the financial crisis, the Fed's asset portfolio was financed primarily by circulating currency, but by October 2014 currency outstanding was about \$1.3 trillion and interest-bearing reserves were about \$2.6 trillion.

The intention of the Federal Open Market Committee (FOMC) was to use the balance sheet expansion as an accommodative tool to supplement its zero interest rate policy, under which the target for the fed funds rate had been reduced to a range of 0–0.25 percent in December 2008. However, the Fed's balance sheet expansion created the necessity for important changes in how the policy directives of the FOMC were implemented in the postfinancial crisis period. Moreover, in ultimately winding down its experiment with unconventional monetary policy—that is, large-scale asset purchases and zero interest rates—the Fed was, and is, sailing in uncharted territory. Therefore, careful evaluation and adjustment along the way was

Cato Journal, Vol. 39, No. 2 (Spring/Summer 2019). Copyright © Cato Institute. All rights reserved.

Stephen D. Williamson is Professor and Stephen A. Jarislowsky Chair in Central Banking at the University of Western Ontario.

and is critical to the Fed's success in managing the experiment and potentially repairing any damage.

What are we to make of the Fed's experiment with a large balance sheet expansion (also known as quantitative easing or QE)? Was the experiment worth it? Has policy implementation been handled correctly during QE? Was the move toward normalization handled in a timely way? What will normalization ultimately entail, and what should it entail? What lessons should we have learned that permit better policymaking in the future?

Fed Intervention before the Financial Crisis

Prior to the large balance sheet period that began in September 2008, the Fed effectively implemented monetary policy in a channel or corridor system. In countries in which interest is paid on reserve balances held with the central bank, channel systems have been formalized as part of central bank communications and implementation. For example, in Canada, the Bank of Canada's policy interest rate is a secured overnight interest rate. The Bank sets a target for that interest rate, and the target falls in a channel. The upper bound on the channel is the interest rate at which the Bank stands ready to lend to financial institutions, which is 25 basis points higher than the target. And the lower bound on the channel is the interest rate at which interest is paid on overnight deposits (reserves) with the Bank, which is 25 basis points below the target. Financial arbitrage dictates that the policy rate must fall between these upper and lower bounds, though typically the Bank achieves the overnight target interest rate with a very small margin of error.

Before the financial crisis, most central banks in rich countries (with some exceptions due to idiosyncratic institutional arrangements) implemented monetary policy in a corridor system that worked similar to Canada's. But monetary policy in the United States worked somewhat differently. First, the Fed targeted an unsecured overnight rate—the fed funds rate. In other countries, unsecured overnight markets may exist, but the target rate—as in Canada—is typically an interest rate on overnight repurchase agreements (repos), that is, a secured rate. Second, the Fed did not pay interest on reserves. This made the effective lower bound on its policy rate zero. However, like other central banks, the Fed conducted lending through the discount window at interest rates higher than the fed

funds interest-rate target, so the fed funds rate was bounded in a channel demarcated by zero on the low side and the discount rate on the high side.

Though the Fed's pre-2008 target interest rate was the unsecured fed funds rate, the Fed did not intervene directly in unsecured credit markets to peg the fed funds rate (nor does it do so currently). In managing its asset portfolio, the Fed focused on an essentially all-Treasury portfolio consisting of bills, notes, and bonds—assets that, for the most part, were held until maturity. Day-to-day intervention to achieve the fed funds rate target occurred in the market for repos. At any given time the Fed was active on both sides of the repo market. That is, it would lend in the repo market, and borrow in terms of reverse repos. Typically, most of the variation in the Fed's repo market intervention occurred through variation in repo activity, rather than reverse repo activity. This intervention procedure is often framed (see Potter 2018) as a process by which the Fed managed the supply of excess reserves, so that the market for excess reserves would clear at an interest rate as close to the fed funds rate target as possible.

It is perhaps more helpful to think of the overnight credit market as involving substitution between secured and unsecured credit. Financial arbitrage between the overnight repo market and the fed funds market is somewhat imperfect because of different timing in these markets during the day (details concerning when the funds go to the borrower one day, and when the debt is settled the next day), friction due to the time it takes to find a counterparty for a particular transaction, and counterparty risk. However, imperfections in arbitrage between secured and unsecured overnight markets did not prevent repo rates from moving together with the fed funds rate. Thus, the Fed's pre-2008 implementation procedure effectively involved influencing repo rates with the goal of pegging the fed funds rate. Sometimes this could be a quite noisy process, particularly during the financial crisis when the fed funds market became contaminated with counterparty risk and the dispersion in interest rates across fed funds transactions became quite large on any given day.

One could certainly make a case that the Fed could have opted for a simpler and more effective channel implementation procedure prior to the financial crisis. Perhaps a better guide for policy is a safe short-term rate of interest—an overnight repo rate rather than the

unsecured fed funds rate—which could have been targeted through a fixed-rate full-allotment auction that would effectively set the overnight repo rate by allowing the quantity traded to vary appropriately. The Fed's focus on the fed funds market seems more the result of historical accident and inertia rather than sound analysis.

Monetary Policy with a Large Balance Sheet

An alternative to a corridor system for monetary policy implementation is a “floor system.” In theory, under such a system the central bank conducts open market operations in such a way that, at market interest rates, there exists a positive stock of overnight excess reserves in the financial system. In general, financial institutions that are permitted to hold reserve accounts will have alternatives to lending to the Fed in the form of excess reserves held overnight. These institutions could also lend overnight on the fed funds market or in the repo market. Financial arbitrage then dictates that, absent any market frictions, the interest on excess reserves (IOER) will determine all overnight market interest rates.

Goodfriend (2002) argued, long before the financial crisis, that a floor system would potentially be a simpler and more effective approach to targeting the fed funds rate in the United States. Rather than intervening indirectly in repo markets to control the fed funds rate—and sometimes missing the target significantly—the Fed could simply pay interest on reserves and implement monetary policy by administering IOER.

In the United States, Congress authorized the payment of interest on reserve balances, to go into effect in 2011. There are standard economic arguments for paying interest on reserves relying on efficiency (going back at least to the work of Friedman 1969), but this legislative change also opened the door to implementation of monetary policy through a floor system. Thus, because of the future asset purchase interventions that the FOMC contemplated in the fall of 2008, payment of interest on reserves was permitted as of October 2008, and the floor system approach became a reality in the United States.

From December 2008 to December 2015, the announced fed funds rate target was 0–0.25 percent, with IOER set at 0.25 percent. In the absence of financial market frictions, theory tells us that the fed funds rate should have been pegged at 0.25 percent by IOER. But, over the December 2008 to December 2015 period, fed funds

typically traded at from 10 to 20 basis points below IOER. Clearly, the floor system did not work in the United States as, for example, Goodfriend (2002) anticipated. This was a cause for concern for the Fed, particularly as regarded the path to policy normalization—the process of ultimately increasing the fed funds target and reducing the balance sheet, possibly to precrisis levels. For example, what would happen if the Fed retained its large balance sheet for some time, while increasing the fed funds rate target? If the Fed were engaged in attempting to hit the increasing target by increasing IOER, would the fed funds rate follow IOER one-for-one, would the margin between IOER and the fed funds rate increase, or would it decrease?

To understand what was going on, it was important to determine the frictions that were behind the difference between IOER and the fed funds rate. Martin et al. (2013) represent conventional views about the source of this interest rate differential. They argued that the interest rate differential was potentially the result of imperfect competition in overnight financial markets. But, perhaps the primary source of the differential, according to them, was bank balance sheet costs, arising due to some quirks in how interest is paid on reserves. That is, when Congress wrote the law governing interest on reserves, it specified that interest could not be paid on the reserve balances of government-sponsored enterprises (GSEs), for example Fannie Mae, Freddie Mac, and the Federal Home Loan Banks. As these financial institutions could potentially be in possession of substantial liquid overnight funds, they had a powerful incentive to earn positive interest on those funds, rather than have them sit in a Fed reserve account earning zero.

But, for private regulated financial institutions holding reserve accounts that bear interest, there are costs to lending on the fed funds market—to GSEs, or to any other financial institutions. That is, deposit insurance premia are tied to bank assets, as are capital requirements. Because a bank's assets and its leverage rise if it borrows on the fed funds markets, there could possibly be substantial balance sheet costs associated with lending on the fed funds market, which could explain a margin of 10 to 20 basis points between IOER and the fed funds rate. Further, these costs could be large enough that the interest rate differential would increase above 20 basis points as IOER went up.

Ultimately, the FOMC decided, prior to “liftoff,” in December 2015, when the fed funds rate target range rose from 0–0.25 percent

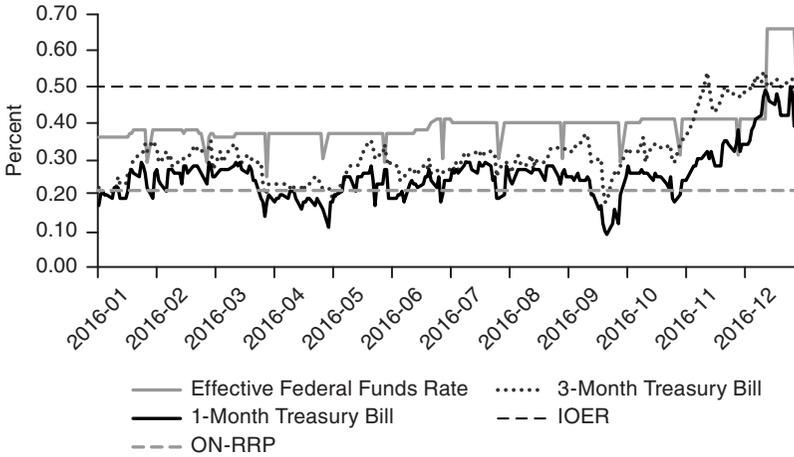
to 0.25–0.50 percent, that the Fed needed to provide a type of sub-floor to the floor system, which had been in place since the end of 2008. In practice, this took the form of an overnight reverse repo (ON-RRP) facility at the New York Fed, which would conduct daily temporary open market operations. As discussed earlier, reverse repos had existed on the Fed's balance sheet prior to the financial crisis, but with the ON-RRP facility, the FOMC envisioned a particular mechanism for maintaining control of the fed funds rate within the announced trading range.

ON RRP's are loans to the Fed (primarily overnight), secured by assets in the Fed's portfolio, with the actual lending conducted through a third party. There is a specified set of counterparties in the ON-RRP market, which include commercial banks and money market mutual funds. It may seem curious that overnight lending to the Fed needs to be secured by collateral posted by the Fed—why would the Fed ever default on an overnight loan? Basically, ON RRP's are reserves by another name, which allow the Fed to extend the reach of its interest-bearing liabilities beyond the financial institutions it traditionally deals with (i.e., those institutions holding reserve accounts) to institutions such as money market mutual funds that cannot hold reserve accounts. ON-RRP lending might also be attractive to GSEs, which do not receive interest on reserve balances, but could hold interest bearing ON-RRP balances with the Fed.

From December 2015 until early in 2018, the Fed set a target range of 25 basis points (beginning with a range of 0.25–0.50 percent) for the fed funds rate, and set IOER at the top of the range, and the ON-RRP rate at the bottom of the range. The ON-RRP rate would then be pegged through a fixed-rate full-allotment procedure, whereby the interest rate is fixed, and take-up in the market is determined by willing lenders at that rate.

Figure 1 shows how the Fed's floor system worked during 2016, when IOER was set at 0.50 percent, and the ON-RRP rate was 0.25 percent. There was significant take-up in the ON-RRP market every day (though perhaps not as large a take-up as was anticipated), and the fed funds rate was typically well within the FOMC's target range. By the end of 2016, the margin between IOER and the fed funds rate had dropped below 10 basis points. Of particular note, as shown in Figure 1, is that 1-month and 3-month Treasury bill interest rates were usually well below the fed funds rate—this in spite of the fact that these T-bill rates include a term premium over overnight

FIGURE 1
FED'S FLOOR SYSTEM, 2016



SOURCE: Federal Reserve Bank of St. Louis, FRED.

lending rates. Thus, in the introductory year of the Fed's floor approach, banks required a premium, which was often more than 25 basis points, to hold reserves rather than 1-month T-bills. Also of note in Figure 1 are the month-end downward spikes in the fed funds rate. This appeared to have been due to end-of-month balance sheet adjustment (for accounting purposes) by key lenders in the fed funds market.

So, in its first year of operation, the Fed's floor system behaved in a manner consistent with conventional understanding of how overnight markets were operating under a large Fed balance sheet. The ON-RRP facility seemed to be important for putting upward pressure on the fed funds rate, and the behavior of the fed funds rate relative to IOER appeared to be consistent with the existence of significant balance sheet costs.

Phasing Out of Reinvestment and Changing Behavior in Overnight Markets

During its balance sheet expansion period from late 2008 until fall 2014, the FOMC established a reinvestment policy, according to which the Fed would replace maturing securities on its balance sheet

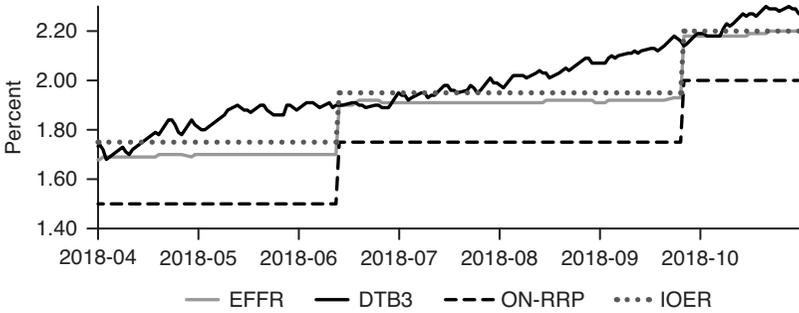
with new asset purchases, so as to maintain a constant nominal balance sheet size even after the cessation in large-scale asset purchases. One option for the Fed would have been to normalize monetary policy by simply reversing the order in which its unusual interventions—during the financial crisis and after—were done. In particular, since the fed funds target was first reduced essentially to zero, followed by three rounds of large-scale asset purchases (QE1, QE2, QE3) and an intervening period of increases in the average maturity of the Fed’s asset portfolio (“operation twist”), why not sell assets to reduce the balance sheet to its former configuration, and then increase the fed funds target?

For the FOMC, the 2013 “taper tantrum” appeared to be a key event. On May 22, 2013, Ben Bernanke announced an imminent—though as yet unofficial—tapering in the Fed’s asset purchase program, then currently underway. The large increase in bond yields that resulted seemed to be unanticipated by the Fed, and perhaps is the primary reason why the FOMC became skittish about balance sheet reduction—or at least outright asset sales as a means to reduce the balance sheet.

By October 2017, after four 25 basis-point increases in the fed funds rate range, the FOMC finally implemented a balance sheet reduction program. This was a rather modest program, entailing caps on the quantities of Treasury securities and mortgage-backed securities that would be permitted to mature within a given month without reinvesting to replace the maturing securities. These caps then increased until reaching their final resting points in October 2018. Since October 2018, the Fed’s securities holdings have been declining, albeit at a slow rate. If the Fed’s assets continue to fall at the current rate, and currency outstanding continues to rise at the current rate, the stock of excess reserves outstanding will reach zero in about 4.5 years.

Coincident with the phasing-out of the Fed’s reinvestment program has been a change in behavior in overnight markets. In Figure 2, we show IOER, the ON-RRP rate, the fed funds rate, and the 3-month Treasury bill rate for the period April 2, 2018 to November 1, 2018. At the March FOMC meeting the fed funds rate range was set at 1.5–1.75 percent, at the June meeting the range changed to 1.75–1.95 percent, and in September this changed to 2.00–2.20 percent. Of particular note is that the difference between IOER and the fed funds rate over this period has fallen from a few

FIGURE 2
SHORT-TERM MARKETS
(APRIL 2, 2018 TO NOVEMBER 1, 2018)



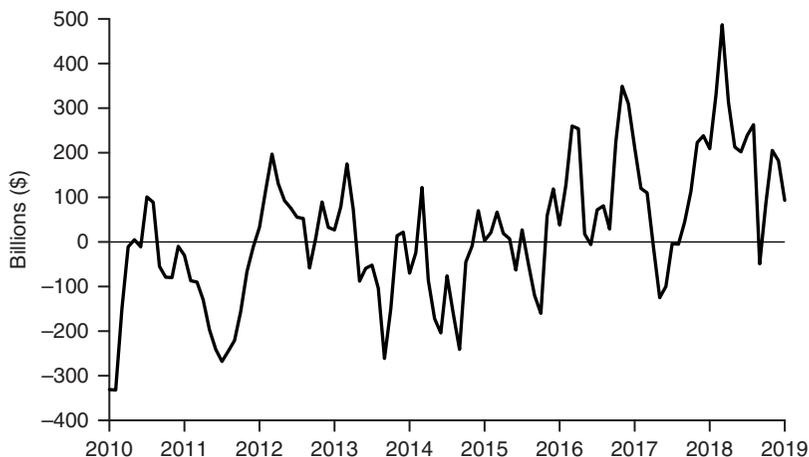
SOURCE: Federal Reserve Bank of St. Louis, FRED.

basis points to zero, with fed funds now trading at IOER. As well, the one-month Treasury bill rate which, as we showed earlier, had fallen below the fed funds rate, and even below the bottom of the fed funds trading range in 2016, is now close to the fed funds rate. Note as well that the downward month-end spikes in the fed funds rate observed during 2016 have disappeared.

Thus, the floor system now appears to be working much more like it does in theory. IOER is now determining overnight rates, and interest rate differentials that formerly appeared highly persistent have gone away. Therefore, if balance sheet costs were important in determining the IOER/fed funds rate differential in 2016, that is no longer the case. Further, the ON-RRP market has recently become essentially inactive, with close to zero take-up in the daily auction. The fact that overnight interest rates, including the fed funds rate, were trading close to the top of the fed funds rate range prompted the FOMC to make an adjustment to the IOER, setting it five basis points below the top of the target range, as of the June 2018 FOMC meeting. The FOMC's goal seems to be to reduce IOER so that the fed funds rate will be roughly in the middle of the target range.

What is going on? The Fed's interpretation (Potter 2018) seems to be that unusually large issues of Treasury bills have encouraged activity in the overnight repo market, so that repo market interest rates have become more competitive relative to fed funds. This has had

FIGURE 3
TREASURY BILL NET ISSUANCE, 6-MONTH MOVING AVERAGE



SOURCE: Securities Industry and Financial Markets Association (SIFMA).

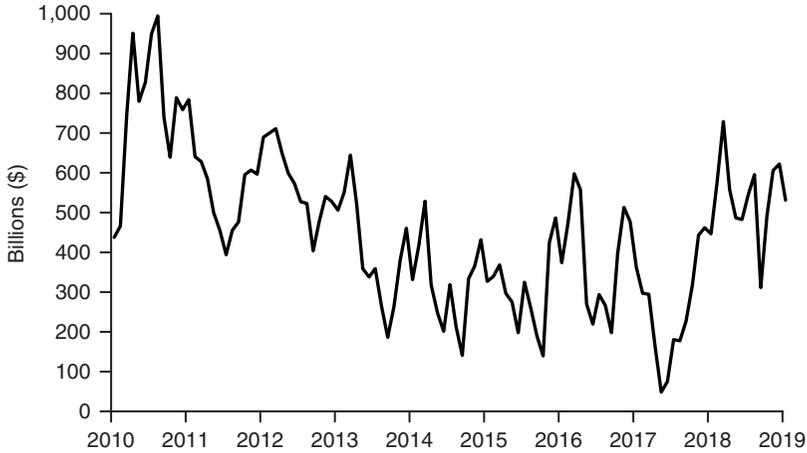
spillovers in the market for Treasury bills as well, with T-bill rates now close to IOER.

Figure 3 shows a six-month moving total of net Treasury bill issuance for the United States, which shows a trend increase in net T-bill issuance in 2018, though the series is quite volatile. However, in Figure 4, total net Treasury issuance, again calculated as a six-month moving total, is higher in 2018 than in the previous two years, but lower than in 2010–12.

So, the story that high net Treasury bill issuance is responsible for the tightening-up of interest rates in the overnight market is questionable. If Treasuries of all maturities are useful as collateral in repo markets, why would there be a substantial difference between IOER and the fed funds rate in 2010–12? And net Treasury bill issuance is highly volatile. So why would a temporary increase in net bill issuance in early 2018 matter, but previous blips in T-bill issuance not matter?

Perhaps, then, the phasing out of the Fed's reinvestment program matters for the behavior we are now seeing in overnight markets. Certainly the timing is hard to ignore, as overnight markets tightened up at the same time reinvestment was being phased out. So, a plausible story is that the phasing out of reinvestment has freed up

FIGURE 4
TOTAL TREASURY ISSUANCE, 6-MONTH MOVING TOTAL



SOURCE: Securities Industry and Financial Markets Association (SIFMA).

on-the-run long Treasury securities for use in overnight repo markets. This, combined with larger net issues of T-bills, increases repo collateral of all maturities, increasing the quantity of credit activity in overnight markets, and putting increased pressure on unsecured fed funds credit.

Does Quantitative Easing Work as Advertised?

Quantitative easing in the form of swaps by the central bank of interest-bearing reserves for long-maturity assets is typically marketed (see, e.g., Bernanke 2010) as a program that exploits segmentation in asset markets. That is, in the markets for Treasury securities, for example, Treasuries of different maturities are imperfectly substitutable, as the argument goes. So, according to the conventional story, even if Treasury bills were essentially perfect substitutes for interest-bearing reserves, if the central bank swaps reserves for long-maturity Treasury securities, in a floor system in which IOER pegs the Treasury bill rate, then long bond yields will fall, as the relative supply of long Treasuries has decreased. Then, according to conventional central bank reasoning, if long bond yields fall, this increases spending and, through a Phillips curve effect, raises inflation.

But, there are good reasons, from theory and evidence, to think that this conventional story does not hold water. First, financial intermediation theory (see Williamson 2017) tells us that central bank intervention matters for economic outcomes because of the special advantages the central bank has over private sector financial intermediaries. In particular, the Bank of England established a model for other central banks in acquiring, over time, a monopoly on the issue of paper currency. Conventional central banking, in a world in which excess reserves are essentially zero, can be viewed as working through asset swaps of outside money—ultimately showing up as currency—for interest-bearing assets (typically government debt). Open market operations matter because private financial intermediaries cannot issue close substitutes for currency.

But, a swap of reserves for long maturity government debt—in a floor system—is a swap of overnight assets for long-maturity government debt, and the private sector seems to be good at converting long-maturity government debt to overnight assets. Both regulated and unregulated financial intermediaries do this. So, a theorist's best guess might be that quantitative easing has no effect at all. Or he or she might go so far as to argue that the Fed is actually worse than private financial intermediaries at converting long-maturity government debt into overnight assets. This is because reserves are a relatively poor overnight asset, as they can be held only by a subset of financial institutions, and those institutions are highly regulated.

There appears to be little or no evidence that QE has any effect on variables that central banks ultimately care about, particularly inflation. Japan is the most obvious case in point in that the Bank of Japan has, since 2013, engaged in a massive quantitative easing program, with the goal of creating a sustained inflation rate of 2 percent. This program has been unsuccessful in that, if we account for the effects of the increase in Japan's consumption tax in 2014, average CPI inflation from 2013 to 2018 has been about zero in Japan.

Evidence that QE may actually be harmful is evident in the observations above on overnight market behavior in the United States after the financial crisis. The effect of the phasing out of reinvestment—the final step in stopping the QE program—has been to, apparently, make overnight markets function more efficiently. Possibly the Fed did harm to financial markets with QE, by replacing good collateral (Treasuries and mortgage-backed securities) with poor collateral (reserves).

What Comes Next?

It appears the FOMC may soon make decisions on the nature of its long-term implementation strategy. The FOMC needs to decide whether it wishes to maintain the current floor system, or revert to mechanisms resembling what existed before the financial crisis. If it keeps the floor system, the key question is how much reserves should be kept in the system so that IOER determines overnight interest rates. Secondary problems are how to communicate policy to the public—as a fed funds rate range, or as a single interest rate, IOER. However, if the Fed reverts to a corridor system, a difference from before the financial crisis will be that reserves pay interest, so the Fed will need to make decisions about the width of the corridor, and perhaps the choice of the target interest rate. For example, the target could be a repo rate rather than the fed funds rate.

It is not clear that the existing floor system has any distinct advantages. First, while a floor system is simple, and accurate in achieving an overnight interest rate target, if the Fed were to target a repo rate in a corridor system, that could also be simple and accurate. Second, it could be argued that an abundance of reserves makes daylight interbank trading more efficient. However, there are approaches that would allow the Fed to advance reserves to financial institutions during the day, and to remove those balances at the end of the day, without implications for overnight markets, within a corridor system. Further, as discussed above, reserves are a poor asset relative to Treasuries, so if a floor system requires a significant quantity of reserves to work, then this implies a significant inefficiency.

Central bankers are concerned about persistently low real interest rates and the implications for monetary policy in the future. If low real interest rates persist, this implies that, to sustain 2 percent inflation, the average short-term nominal interest rate must be lower than in the past. Central bankers speculate that this implies that central bankers will encounter the effective lower bound on nominal interest rates with higher frequency in the future than was the case in the past. At the effective lower bound, central banks—including the Fed—will be tempted to resort to balance sheet expansions. As has been discussed, it seems hard to make a case that this would be a good idea. Unfortunately, central bankers have a difficult time admitting errors, which increases the chances of repeating those errors.

References

- Bernanke, B. (2010) “The Economic Outlook and Monetary Policy.” Speech at the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyo. (August 27).
- Friedman, M. (1969) “The Optimum Quantity of Money.” In *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine.
- Goodfriend, M. (2002) “Interest on Reserves and Monetary Policy.” Federal Reserve Bank of New York *Policy Review* 8 (1).
- Martin, A.; McAndrews, J.; Palida, A.; and Skeie, D. (2013) “Federal Reserve Tools for Managing Rates and Reserves.” Federal Reserve Bank of New York Staff Report No. 642.
- Potter, S. (2018) “U.S. Monetary Policy Normalization Is Proceeding Smoothly.” Speech at the China Finance 40 Forum–Euro 50 Group–CIGI Roundtable, Banque de France, Paris, France (October 26). Available at www.fedinprint.org/items/fednsp/297.html.
- Williamson, S. (2017) “Quantitative Easing: How Well Does This Tool Work?” Federal Reserve Bank of St. Louis *Regional Economist* (3rd quarter): 8–14.