FLOORED!
How a Misguided Fed Experiment Deepened and Prolonged the Great Recession and Why the Fed – or Congress – Ought to End It

By George Selgin
March 1, 2018
FLOORED!
How a Misguided Fed Experiment
Deepened and Prolonged
the Great Recession
and Why the Fed
- or Congress -
Ought to End It

George Selgin
Director
Center for Monetary and Financial Alternatives
The Cato Institute
gselgin@cato.org

February 26, 2018

Preliminary. Comments are welcome. The author thanks David Beckworth, Charles Calomiris, Donald Dutkowsky, Charles Goodhart, David Laidler, Lydia Mashburn, Norbert Michel, David VanHoose, and Stephen Williamson for their helpful criticisms and suggestions.
# CONTENTS

I. Introduction .................................................................................................................. 1

II. Prelude: Monetary Control Before the Crisis ................................................................. 3
   II. a. Targeting the Fed Funds Rate .............................................................................. 3
   II. b. Open Market Operations .................................................................................... 4
   II. c. The Monetary “Transmission Mechanism.” ......................................................... 6

III. Interest on Bank Reserves .......................................................................................... 9

IV. The Floor System’s Beginnings .................................................................................. 13
   IV. a. Fear of Falling .................................................................................................... 13
   IV. b. From Corridor to Floor ..................................................................................... 15
   IV. c. A Floor with a Sub-floor .................................................................................. 18
   IV. d. A Banking System Tipping Point ..................................................................... 19
   IV. e. The Floor System and the Friedman Rule ......................................................... 22
   IV. f. A Dubious Advantage ....................................................................................... 23

V. Digression: Is the Floor System Legal? .................................................................... 27
   V. a. The Law’s Original Intent: A Below-Market IOER Rate ...................................... 27
   V. b. Above the Law? .................................................................................................. 28
   V. c. “One of these rates is not like the others...” ...................................................... 30

VI. The Floor System and Interbank Lending .................................................................. 32
   VI. a. Goodbye Fed Funds Market .............................................................................. 32
   VI. b. IOER vs. Perceived Counterparty Risk .............................................................. 34
   VI. c. From Lender to Borrower of First Resort .......................................................... 35

VII. The Floor System and Reserve Hoarding ................................................................. 37
    VII. a. The Accumulation and Distribution of Excess Reserves .............................. 37
    VII. b. Excess Reserves and the Fed’s Balance Sheet .............................................. 43
    VII. c. Excess Reserves and the Liquidity Coverage Ratio ....................................... 46
    VII. d. Reserve Demand and Opportunity Cost ..................................................... 48

VIII. The Floor System and Retail Lending ................................................................... 50
    VIII. a. Lending Before and Since the Crisis ............................................................ 50
    VIII. b. The Direct Influence of IOER on Bank Lending ......................................... 51
I. Introduction

Federal Reserve authorities responded to the 2007-8 financial crisis with a sequence of controversial monetary policy experiments aimed at containing the crisis and, later on, at promoting recovery. One of those experiments consisted of the Fed’s decision to start paying interest on depository institutions’ balances with it, including both their legally required balances and any balances they held in excess of legal requirements. Because the interest rate on excess reserves was high relative to short-term market rates, the new policy led to the establishment of a “floor”-type operating system, meaning one in which changes in the rate of interest paid on excess reserves, rather than open-market operations, became the Fed’s chief instrument of monetary control.

Although it has attracted less attention, and generated less controversy, than many of the Fed’s other crisis-related innovations, the Fed’s shift to a floor system has also had more profound and enduring consequences than many of them. And despite Fed officials’ intentions, those consequences, including a radical change in the Fed’s methods of monetary control, have mostly been regrettable. While Fed officials hoped that the new floor system would assist them in regulating the flow of private credit in the face of extremely low and falling interest rates, a close look at the workings of the system, and at its record, shows that those hopes have been disappointed.

As the following pages will show, among its other consequences the Fed’s new operating system

- intensified an already severe economic downturn by serving as the means by which the Fed maintained an excessively tight monetary policy;
- led to a sustained collapse in the interbank market for federal funds, thereby destroying the Fed’s traditional means of monetary control;
- dramatically reduced the effectiveness of open-market operations, so that even massive Fed asset purchases might not supply the stimulus to investment and spending that much smaller purchases would once have achieved;
• undermined productivity by substantially increasing the Fed’s role in allocating scarce credit; and
• made it more difficult for the Fed to reach its 2 percent inflation target.

Although the FOMC announced in 2015 that it would “begin an extended effort to evaluate potential long-run monetary policy implementation frameworks” which was “expected to run through the end of 2016” (Board of Governors 2015), if it has reached any conclusions based on that effort, it has yet to announce them. And despite having finally embarked upon a program of monetary policy “normalization,” involving the gradual lifting of its policy rates and a reduction in the size of its balance sheet, many top Fed officials appear to be strongly inclined to make the floor system permanent.

Those officials claim, nevertheless, to be willing to entertain the possibility of switching from that system to a more conventional “corridor”-type system in the future. This book’s purpose is to encourage them in that direction, while also making others aware of the present system’s shortcomings. It will explain in detail how the Fed’s floor-system experiment came about, what its intended and actual consequences have been, and why either the Fed itself or Congress should bring the experiment to an end as rapidly as can be done without causing further economic damage.
II. Prelude: Monetary Control Before the Crisis

To appreciate the radical changes made to the Fed’s operating system during the financial crisis, it’s helpful to first review how the Fed managed monetary policy before then, and particularly during the period known as the “Great Moderation,” a time of modest and stable inflation and otherwise relatively low macroeconomic volatility that began in the mid-1980s and ended when the crisis struck.

II. a. Targeting the Fed Funds Rate

It happens that the Great Moderation period almost perfectly coincided with the FOMC’s decision, following Paul Volcker’s successful taming of the Great Inflation of the preceding decade, to implement monetary policy by setting an explicit target for the federal funds rate—the overnight interest rate at which institutions that keep deposit balances at the Fed lend such balances to, or borrow them from, other such institutions. The fed funds market served in those days mainly as a means by which banks in danger of falling short of their minimum reserve requirements could make up for reserve shortages by borrowing from banks with reserves to spare. Because bank reserves didn’t bear any interest, and the fed funds rate was always positive, banks with surplus or “excess” reserves were always happy to lend those reserves overnight. Thanks to the fed funds market, banks could collectively make do with very few excess reserves, knowing that interbank lending would place available reserves wherever they were needed.

Back then, as ever since, the Fed operated under the so-called “dual mandate”—the requirement, first incorporated into the Federal Reserve Act in 1977, that the Fed manage monetary policy so as to "promote effectively the goals of maximum employment, stable prices, and moderate long term interest rates." Under the fed funds rate targeting regime, the Fed faced a twofold challenge. First of all, it had to choose a fed funds rate target that it hoped would prove consistent with fulfilling the dual mandate, meaning a target that would be low enough to guard against excessive unemployment,

---

1 Although the Fed allows banks to overdraw their accounts during the day, it requires them to end each day with non-zero balances sufficient to meet their reserve requirements, or else pay a penalty.
while high enough to keep inflation under control. Then it had to manage the supply of bank reserves so as to keep the actual federal funds rate at its targeted level.

The details concerning just how the FOMC decided where to set its fed funds target need not concern us. Instead, we may simply observe that, although the FOMC’s choices were supported by rafts of statistics and elaborate forecasts based upon them, they also involved a measure of trial and error. If, for example, the FOMC determined that the Fed’s policy stance had allowed, or was likely to allow, an unwanted decline in inflation, or increased unemployment, or a combination of both, it would adopt a looser policy stance, which meant lowering its fed funds rate target—a lower target meaning a loosened market for bank reserves, and a general inclination for banks to loosen their own lending terms. If, on the other hand, the likely course of inflation or unemployment or both suggested the need for a tighter policy stance, the FOMC would raise its fed funds rate target, encouraging banks to tighten in turn.

Although the FOMC never relied on any simple algorithm to determine its rate target, John Taylor (1993) found that its rate settings during much of the Great Moderation period\(^2\) could be approximated by the simple formula, now known as the “Taylor Rule,”

$$\text{FFR}^* = 2 + \pi_t + .5(\pi_t - \pi^*) + .5(y_t - y_{t^*}),$$

where FFR\(^*\) is the chosen rate target, 2 is the assumed, constant long-run value of the real (inflation adjusted) fed funds rate, \(\pi_t\) and \(\pi^*\) are the actual and target inflation rate, and \(y_t\) and \(y_{t^*}\) are actual and “potential” (that is, “maximum employment”) levels of output.

II. b. Open Market Operations

Importantly, the federal funds rate whose value the FOMC endeavored to control was a private-market rate, the actual level of which depended, like the values of all

\(^2\) Taylor (2009, chapter 1) has since claimed that the Fed’s policy stance after the dot-com crash of 2001 was easier than what the Taylor Rule would have prescribed, and that by departing from that rule the Fed inadvertently contributed to the subprime boom and bust.
market-determined interest rates, on the interaction of supply and demand—specifically, the supply of and demand for reserve balances at the Fed, aka "federal funds."

Because the Fed did not pay interest on its deposit balances in those days, the only direct influence it had on banks’ demand for federal funds was through its setting of minimum reserve requirements. But while these requirements were occasionally adjusted over the course of the Great Moderation, the adjustments were aimed, not at influencing the stance of monetary policy, but at relaxing the burden the requirements placed on banks subject to them (Feinman 1993). So far as the Fed’s day-to-day operations were concerned, monetary control was a matter of adjusting the supply of federal funds so as to make the funds rate land on target.

To do that, the Fed relied upon “Open Market Operations,” meaning purchases or sales of government securities, and short-term Treasury securities especially, from a score or so of approved (“primary”) security dealers. To add to the supply of federal funds, and thereby put downward pressure on the fed funds rate, the Fed would buy securities; to reduce the supply, it would sell securities in its possession. The operations were handled by the New York Fed’s Open Market Trading Desk, and supervised by the FOMC-appointed manager of its System Open Market Account (SOMA). The Desk would arrange the necessary auctions, which were usually held daily, based on estimates of the direction and scope of the operations needed to move the funds rate to its target, or to keep it from moving away from it.

How did open-market operations alter the supply of federal funds? When the Fed bought securities, it paid dealers who made the sales by crediting the dealers’ own Fed deposit balances (if the dealers were themselves banks) or by crediting the balances of the dealers’ banks, and so increased the total supply of federal funds by the amount of the purchase. When it sold securities, it debited dealer and dealer-affiliated reserve balances by the amount of their successful bids, reducing total quantity of outstanding federal funds by the same amount.
Because keeping the actual fed funds rate near its target often meant adjusting the supply of reserves to meet temporary rather than persistent changes in the demand for them, the Fed undertook both "permanent" and "temporary" open-market operations, where the former involve "outright" security purchases or (more rarely) sales, and the latter involve purchases or sales accompanied by "repurchase agreements" or "repos." (For convenience's sake, the term "repo" is in practice used to describe a complete sale and repurchase transaction.) For example, the Fed might purchase securities from dealers on the condition that they agreed to repurchase those securities a day later, thereby increasing the supply of reserves for a single night only. (The opposite operation, where the Fed sells securities with the understanding that it will buy them back the next day, is called an "overnight reverse repo.") Practically speaking, repos are collateralized loans, except in name, where the securities being purchased are the loan collateral, and the difference between their purchase and their repurchase price constitutes the interest on the loan which, expressed in annual percentage terms, is known as the "repo rate." The obvious advantage of repos, and of shorter-term repos especially, is that, because they are self-reversing, a central bank that relies extensively on them can for the most part avoid resorting to open-market sales when it wishes to reduce the supply of federal funds. Instead, it merely has to refrain from "rolling over" or otherwise replacing some of its maturing repos.

II. c. The Monetary “Transmission Mechanism.”

Although the immediate goal of the Fed’s open-market operations was to keep the federal funds rate on target, their ultimate purpose was achieving a monetary policy stance consistent with the Fed’s dual mandate. As we’ve seen, the Taylor Rule supplies a rough indication of how the Fed adjusted its fed funds rate target in response to actual and desired levels of inflation and output. The Great Moderation suggests, furthermore, that by setting its fed fund targets as if it were heeding that rule, the Fed was in fact able to keep actual inflation and output from veering far from their desired levels.
The causal chain connecting the Fed’s rate targeting procedures and its ultimate policy objectives is sometimes referred to as the “monetary transmission mechanism.” To consider a very simple case, albeit one that’s especially pertinent to our subject, suppose that, at its current rate setting, the FOMC anticipates below target inflation and unemployment. Consequently, it reduces its target setting, while instructing the SOMA account manager to arrange open-market security purchases sufficient to drive the actual funds rate down to the newly-reduced target level. The purchases—say they end up amounting to $20 billion—increase banks’ reserve balances by a like amount. Assuming that the banks had been meeting their reserve requirements, this also means that they find themselves collectively holding $20 billion in excess reserves. Because bank reserves earn no interest, banks seek to dispose of the excess, initially by offering them in the fed funds market. That pushed the fed funds rate down, helping the Fed to hit its lowered target.

But that’s not the whole story, because lending on the fed funds market only shifts excess reserves around, without reducing the total quantity. As banks (again, collectively) find themselves confronted with a persistent reserve surplus, they’ll engage in other sorts of lending, including retail lending, which involves crediting borrowers’ deposit accounts. Such lending will continue until bank deposits have grown enough to transform the former excess reserves into required reserves, achieving a full equilibrium of reserve supply and demand. The expansion of bank loans and deposits, each by some multiple of the new reserves, contributes to a like expansion of total spending, which puts upward pressure on prices (or at least combats any tendency for them to decline), while reducing slack in the labor market. Thus, besides serving to keep the fed funds rate on target, the Fed’s open-market operations ultimately serve to keep inflation and employment themselves at their desired levels.

***

Such, in a nutshell, was monetary control prior to the Subprime Crisis. The rest of this book is about how a seemingly minor change in Fed policy overturned the old system
of monetary control, replacing it with one that was dramatically different—and far inferior.
III. Interest on Bank Reserves

Economists have long understood that, so long as bank reserves bear no interest, minimum reserve requirements, whether they're met using banks’ holdings of vault cash or with Federal Reserve deposit balances, act like a tax on bank deposits, and therefore on bank depositors. Although the Fed earns interest on the assets backing banks’ reserves, until October 2008 it didn’t share that interest with commercial banks. Instead, it remitted all its interest earnings, net of its operating expenses, to the U.S. Treasury.

Though it was only in the midst of the recent financial crisis that the Fed first began paying interest on bank reserves, the possibility of its doing so has long been a subject of discussion and debate. Indeed, the idea was initially broached during the discussions that led to the passage of the original Federal Reserve Act in 1913. That original suggestion was ultimately rejected, in large part because of opposition from Wall Street banks, which saw it as a threat to their lucrative correspondent business.3

So matters stood for more than half a century, thanks to the generally low inflation and interest-rate environment that prevailed during most of that time, and, after 1933, to the fact that Regulation Q and other provisions of the 1933 Banking Act relieved commercial banks themselves of pressure to pay competitive rates of interest on their own deposit balances.

Starting in the mid 1960s, however, a combination of rising inflation, declining Fed membership, the rise of Money Market Mutual Funds, and increasingly intense global banking competition, revived Fed officials’ desire to be able to pay interest on bank reserves, as an alternative to dispensing with mandatory reserve requirements, which they regarded as an aid to monetary control (see Weiner 1985; Higgins 1977; and Eubanks 2002). Over the course of the next several decades Fed officials tried several times to gain Congress’s permission to pay interest on reserves.4 Until 2006 these attempts were

---

3 On Wall Street banks’ role in shaping the Federal Reserve Act, see Selgin (2016).

4
successfully opposed by the U.S. Treasury, which feared having its seigniorage earnings substantially reduced. But in that year the Fed finally managed to have the authority it had long sought included among the provisions of the Financial Services Regulatory Relief Act.

The Fed’s ultimate success was made possible in large part by reduced Treasury opposition, itself due to a considerable decline, during the 1990s, in the burden posed by mandatory reserve requirements, and the corresponding decline in the Treasury’s seigniorage revenues. Although actual requirements were reduced somewhat, their reduced burden was mainly a result of banks’ successful employment of “sweep accounts” to avoid them. By substantially reducing the effective reserve tax base, these developments also reduced the cost to the Treasury of allowing the Fed to pay interest on reserves.

By the same token, however, the reduced burden of reserve requirements also limited the “regulatory relief” banks would gain from interest payments on reserves. Perhaps in recognition of this, Fed officials, in making their successful bid for the right to pay interest on bank reserves, offered new grounds for doing so that had nothing to do with reducing the reserve tax. In particular, then Fed Governor Donald Kohn (2005) argued that, besides making it unnecessary for banks to resort to sweep accounts and other “reserve avoidance measures,” paying interest on reserves would assist the Fed in conducting monetary policy “by establishing a sufficient and predictable demand for balances at the Reserve Banks so that the System knows the volume of reserves to supply

---

4 As Goodfriend and Hargraves (1983, pp. 16–17) report, in 1978 the Fed went so far as to declare that, because statute law didn’t expressly prohibit it from doing so, it planned to start paying interest on reserves without Congress’s permission. That gambit came to grief when Representative Henry Reuss and Senator William Proxmire, the chairmen of the House and Senate Banking Committees, respectively, called it “a blatant usurpation of Congressional powers [that] would raise profound questions about the continued independence of the Fed.”

5 Sweep accounts are actually pairs of accounts set up for the same account balance so that funds can be shifted or “swept” automatically from one account to the other. To avoid reserve requirements, during overnight reserve monitoring periods banks sweep customers’ balances from accounts subject to such requirements, such as checking accounts, into exempt ones, such as money market accounts.
(or remove) through open market operations to achieve the FOMC’s target federal funds rate."

Importantly, in view of later developments, Kohn’s statement implied that interest on reserves was meant to support, rather than supplant, the Fed’s traditional methods of monetary control, including its reliance upon open-market operations as its chief tool for reaching its monetary targets. Interest on reserves, Kohn said,

Would act as a minimum for overnight interest rates, because banks would not generally lend to other banks at a lower rate than they could earn by keeping their excess funds at a Reserve Bank. Although the Board sees no need to pay interest on excess reserves in the near future, and any movement in this direction would need further study, the ability to do so would be a potentially useful addition to the monetary toolkit of the Federal Reserve (ibid; emphasis added).

By recommending that the Fed set aside the possibility of paying interest on excess reserves for “further study,” Kohn makes clear that his own case for paying interest on reserves is one for paying interest on required reserves only. As events would show, paying interest on excess reserves could actually undermine both efficiency and monetary control, by causing banks to accumulate unlimited quantities of excess reserves.

Kohn’s remarks also make it clear that, far from even entertaining a radical change in its operating system, the Fed was merely toying with the possibility of eventually paying interest on banks’ excess reserve balances for the purpose of establishing an orthodox “corridor” system of the sort that several other central banks were then employing. In an orthodox corridor system, a variable but generally positive interest rate on bank reserves, rather than a zero rate, serves as a lower bound for the central bank’s policy rate, while the central bank’s emergency lending rate serves as an upper bound. Although the policy rate can vary within these limits, it generally stays close to a target
set, in the most common “symmetrical” corridor arrangements, half-way between them.\(^6\)

To keep the policy rate on target, the central bank relies on a combination of open-market operations and changes to the interest rate on reserves, where the first alters the supply of, and the last the demand for, overnight funds. However, except on those infrequent occasions when either of the corridor’s limits becomes binding, open-market operations continue to serve as the central bank’s chief monetary policy instrument (Kahn 2010, pp. 13–15). For that reason a switch to a corridor system would not have involved any radical change in the Fed’s monetary control procedures.

Indeed, the Fed’s “Great Moderation” operating system, discussed earlier, might be described as a corridor system of sorts, albeit one that involved an asymmetrical and variable corridor based on a fixed, zero IOER rate and a discount (“primary credit”) rate set at a fixed spread above the federal funds rate target.\(^7\) As we’ve seen, in that regime the Fed relied on open-market operations to achieve its policy target. Had it actually employed interest on reserves to establish a proper corridor system, as it planned to do in 2006, and even had it allowed interest to be paid on excess reserves with that aim alone in mind, paying interest on reserves wouldn’t have constituted a radical change. But as we shall see, when the Fed actually put its new tool to work, a corridor system was no longer what it had in mind.

---

\(^6\) Before 2008, symmetrical corridor systems with positive deposit rates were employed by the ECB, the Bank of England, the Riksbank, the Bank of Canada, the Reserve Bank of Australia, and the Reserve Bank of New Zealand, among other central banks, while Norges Bank maintained an asymmetrical corridor.

\(^7\) The primary credit-fed funds target rate spread, which was set at 100 basis points from January 9, 2003 until August 17, 2007, was halved on the latter date, and halved again, to just 25 basis points, on March 16, 2008. On February 19, 2010, the spread was returned to 50 basis points, where it has since remained.
IV. The Floor System’s Beginnings

IV. a. Fear of Falling

The 2006 Financial Services Regulatory Relief Act would have allowed the Fed to begin paying interest on depository institutions’ reserve balances commencing October 1, 2011. However, the worsening financial crisis of 2008 led to the passage of the Emergency Economic Stabilization Act, which advanced the effective date of the 2006 measure to October 1, 2008.

Fed officials sought and received Congress’s authorization to begin paying interest on reserves three years ahead of the originally planned date for a reason completely unrelated to those that Kohn and others had offered in defense of the original measure. As Ben Bernanke explains in his memoir,

We had initially asked to pay interest on reserves for technical reasons. But in 2008, we needed the authority to solve an increasingly serious problem: the risk that our emergency lending, which had the side effect of increasing bank reserves, would lead short-term interest rates to fall below our federal funds target and thereby cause us to lose control of monetary policy. When banks have lots of reserves, they have less need to borrow from each other, which pushes down the interest rate on that borrowing—the federal funds rate.

Until this point we had been selling Treasury securities we owned to offset the effect of our [emergency] lending on reserves (the process called sterilization). But as our lending increased, that stopgap response would at some point no longer be possible because we would run out of Treasuries to sell. At that point, without legislative action, we would be forced to either limit the size of our interventions...or lose the ability to control the federal funds rate, the main instrument of monetary policy. [By] setting the interest rate we paid on reserves high enough, we could prevent the federal funds rate from falling too low, no matter how much [emergency] lending we did (Bernanke 2015, pp. 325–6; emphasis added).
The same understanding of the Fed’s intention in implementing IOER three years ahead of the original, 2006 schedule was conveyed in the Board of Governors’ (2008a) October 6, 2008 press release announcing the Fed’s new tool:

The payment of interest on excess reserves will permit the Federal Reserve to expand its balance sheet as necessary to provide the liquidity necessary to support financial stability while implementing the monetary policy that is appropriate in light of the System’s macroeconomic objectives of maximum employment and price stability.

More than a year later, with the advantage of hindsight, Richmond Fed economists John R. Walter and Renee Courtois (2009) offered an almost identical account. The Fed’s emergency credit injections, they wrote,

had the potential to push the fed funds rate below its target, increasing the overall supply of credit to the economy beyond a level consistent with the Fed’s macroeconomic policy goals, particularly concerning price stability. ... Once banks began earning interest on the excess reserves they held, they would be more willing to hold on to excess reserves instead of attempting to purge them from their balance sheets...

Although their meaning may seem surprising in light of subsequent developments, these official statements, as well as many others like them, make it clear that the Fed’s main concern in October 2008 was that of avoiding an unplanned *loosening* of what it still considered an appropriate monetary policy stance. Although they were keen on providing emergency support to particular firms and markets, Fed officials recognized no general liquidity shortage calling for further monetary accommodation. The challenge, as they saw it, was that of extending credit to particular recipients without letting that credit result in any general increase in lending and spending.

Figure IV.1 (below) may further clarify the Fed’s reasoning. The solid line in it shows the Fed’s total assets, while the dashed line shows its Treasury holdings, before and
since Lehman’s failure. That failure was followed by a dramatic increase in the Fed’s emergency lending. But because the Fed’s Treasury holdings had already fallen by then to what Fed officials considered a minimal level, and because those officials were determined to keep the fed funds rate from falling below its target, they needed some other way to keep new reserves the Fed was creating from flooding into the fed funds market. While the Treasury, at the Fed’s behest, did its part by diverting funds to a “Supplementary Finance Account” created for the purpose of reducing banks’ share of total Fed balances (dotted line), for the most part the Fed was counting on IOER to encourage banks to accumulate excess reserves instead of lending them.

![Figure IV.1: Federal Reserve System Assets and Treasury Supplementary Finance Account](image)

**IV. b. From Corridor to Floor**

Fed officials’ desire to treat IOER as a substitute for offsetting the Fed’s loans with bond sales was fundamentally at odds with having the IOER rate serve as the lower-bound of a corridor system, orthodox or otherwise. Yet the previously-mentioned press release suggests that they were at first unaware of this. “The rate paid on excess balances,” the release announced, “will be set initially as the lowest targeted federal funds rate for each reserve maintenance period less 75 basis points,” which setting would “establish a lower bound on the federal funds rate.” The intent here seemed to be that of establishing an orthodox corridor system, in which the IOER rate is below the monetary authority’s intended policy rate target, with changes in the stock of bank reserves serving to keep the rate near that target.
But to serve as a means for preventing banks from disposing of their excess reserves in the fed funds market, and from thereby undermining the Fed’s attempts to keep the fed funds rate on target, the IOER had to be at least as high as, and probably higher than, prevailing rates on other short-term loans, including the overnight interbank rate. Just how an IOER rate set 75 basis points below “the lowest targeted federal funds rate” could do that, and specifically how it could keep the effective fed funds rate from eventually slipping as much as 75 basis points below the Fed’s stated target, the Board’s press release didn’t explain. Nor could it have, since IOER could only keep the fed funds rate from falling below the Fed’s target if the IOER rate was set equal to, or rather (for reasons we’ll come to) above, the target. Partly for this very reason, the effective fed funds rate continued to decline.

What the Fed actually needed, if it was to maintain its desired fed funds target, was not a corridor system, but what Marvin Goodfriend (2002) and others have called a “floor” system. In a floor system the IOER rate itself becomes the central bank’s policy rate, and hence its chief instrument of monetary control, replacing management of the stock of bank reserves in that role. The difference between the two arrangements is illustrated in Figure IV.2. In a corridor system, as we’ve seen, the target fed funds rate is set between, and typically half-way between, the IOER rate and the discount (or primary credit) rate, and either open-market operations or changes in the interest rate on excess reserves can be employed to keep the effective funds rate close to its target value. In a floor system, in contrast, the Fed sets an at or above-market IOER rate equal to its desired fed funds rate target, thereby allowing the IOER rate to serve, in Goodfriend’s words, as both a “floor under which banks would not lend reserves to each other” and “a ceiling above which banks would not lend to each other.” Because banks can earn at least as

---

8 In a conference call held on September 29, 2008, or just days before the Emergency Economic Stabilization Act was passed, the FOMC, which was then planning to set the IOER rate at 50 points below its fed funds target, was told by Brian Madigan, the FOMC’s Secretary and Director of the Fed’s Division of Monetary Affairs, that the arrangement they were contemplating “may turn out [to] operate more like what we have been calling a floor system.” However, Madigan added, while the FOMC might eventually “need to use the excess reserve rate as the way to effectively set the federal funds rate,” it would have “to see how it goes and get some experience” (Board of Governors 2008a).
much by holding reserve balances (“federal funds”) as they can by lending them, and with
less effort, the demand schedule for fed funds becomes flat – that is, infinitely interest-
elastic – at the IOER rate. The Fed could then maintain a desired fed funds rate target
even as it flooded the market with bank reserves.

![A Corridor System](image1)
![A Floor System](image2)

Figure IV.2: Corridor and Floor Operating Systems
(Source: Reproduced from Keister 2012)

As it happens, the Board’s press announcement provided, conveniently, that “the
formula for the interest rate on excess balances may be adjusted subsequently in light of
experience and evolving market conditions.” The Fed was, unsurprisingly, quick to take
advantage of this clause. The Fed’s initial IOER rate settings, shown in Figure IV.4, saw it
stumble during the new regimes’ first month from its original, corridor-style rate setting
plan to a floor-system plan consistent with its goal of encouraging banks to accumulate,
rather than lend, federal funds. Having initially set the IOER rate at 75 basis points, that
is, 75 basis points below its 150 basis point federal funds rate target, and also below the
then-prevailing effective federal funds rate, the Board raised it to 115 basis points on
October 23rd, narrowing the IOER-fed funds target spread to just 35 basis points. It
maintained that spread at first as it lowered the funds rate target to 100 basis points on
October 29th. But on November 6th, just one month after having announced its new
policy, it eliminated the spread entirely by raising the IOER rate to 100 basis points. From
that point onward, the effective fed funds rate remained persistently below the IOER rate,
notwithstanding the Fed’s December 17, 2008 decision to reduce the IOER rate to just 25 basis points, where it stayed for seven full years. A floor system was thus established, albeit one with some unorthodox properties.

Although the Fed’s IOER adjustments achieved its immediate purpose of discouraging banks from lending their excess reserves on the fed funds market, they failed to achieve its ultimate goal of keeping the effective fed funds rate on target. Instead, as the Fed made its way from a corridor to a floor system, the gap between the Fed’s rate target and the effective fed funds rate continued to widen. By November 6th the effective fed funds rate had plunged to just 23 basis points—far below the Fed’s 100 basis point target.

That IOER failed to keep the fed funds rate on target even once the IOER rate was no longer set below that target was both inconsistent with the way a floor system was supposed to operate, and a source of considerable disappointment to Fed officials and staff economists. Blame for it has been placed on the fact that, in addition to banks, various GSEs, including Fannie Mae, Freddy Mac, and the Federal Home Loan Banks, keep deposit balances at the Fed, but aren’t eligible for interest on those balances.\(^9\) The

\(^9\) See Kahn (2010).
GSEs access to the fed funds market therefore created an arbitrage opportunity Fed officials didn’t anticipate, with GSEs lending fed funds overnight to banks, and especially to very large U.S. banks and a handful of (mostly European) foreign bank branches, in exchange for a share of the latter’s IOER earnings. When the IOER rate was 25 basis points, for example, one of the Federal Home Loan Banks might lend funds overnight to a commercial bank for less than 25 basis points, allowing the commercial bank to profit from the spread, while securing for itself a return greater than the zero rate it would earn if it just held on to its Fed balance.

Consequently, instead of getting the solid floor it wanted, the Fed had to settle for a “leaky” floor. Indeed, because the effective fed funds rate tended to fall below the IOER rate, the latter ended up looking less like a floor than like a ceiling—just the opposite of corridor arrangement. When, in mid-December 2015, the Fed began making use of a new overnight reverse repurchase agreement (ON-RRP) facility to establish what Stephen Williamson (2016) has called a “floor-with-sub-floor” system, with the effective fed funds rate trading within a target “range” defined by the IOER rate (floor), and the ON-RRP (subfloor), the resemblance of the Fed’s new system to a corridor system gone topsy-turvy became all the more complete.

**IV. d. A Banking System Tipping Point**

It cannot be stressed enough that, although the interest rate paid on excess reserves seemed paltry, and especially so between December 2008 and December 2015, when it was just 25 basis points, that paltry rate was nevertheless sufficiently high, relative to other lending rates—and to overnight lending rates especially—to inspire the dramatic changes in bank behavior that are the necessary counterpart of a switch from a corridor-type to a floor-type central bank operating framework.

In particular, as David VanHoose and Donald Dutkowsky (2017) convincingly demonstrate, even at its lowest level the Fed’s IOER rate was indeed high enough to cause banks to respond to the Fed’s reserve creating activities by accumulating excess reserves instead of taking part, as they otherwise would have, in additional wholesale lending.
Using a static bank optimization model, VanHoose and Dutkowsky first show that four distinct equilibrium outcomes are possible, depending on the value of various interest rate and bank resource cost parameters, and also on the banking system’s required reserve ratio. These outcomes are (1) a “corridor” regime in which banks do not hold excess reserves, but do engage in wholesale lending; (2) a “floor” regime in which they accumulate excess reserves, but do not engage in wholesale lending; (3) a “mixed” regime in which they engage in wholesale lending but also accumulate excess reserves and (4) a “neither” regime in which they refrain from both wholesale lending and reserve accumulation, devoting their resources exclusively to retail lending.

VanHoose and Dutkowsky find, first of all, that their model’s “mixed” and “neither” regime outcomes occur only for relatively few parameter combinations, with most combinations leading to either a corridor- or a floor-system outcome. Furthermore, only very modest changes in the IOER rate may suffice to trigger a switch from a corridor to a floor regime, and especially so when other interest rates are relatively low. VanHoose and Dutkowsky then go on to test their model using calibrated values of its structural parameters for the pre- and post-October 2008 periods. They find that the model does indeed predict the regime change that actually occurred, and that that change can in fact be attributed to the Fed’s decision to begin paying interest on bank reserves.

Figure IV.4, reproduced from VanHoose and Dutkowsky (2017, p. 11), shows the regimes that will prevail for various combinations of the IOER rate (rQ) and the wholesale lending rate (rF), given other assumed parameter values, and the boundary lines separating them. For all combinations above the solid boundary line, banks refrained from wholesale lending (F = 0); for all combinations below the dashed line, they refrain from accumulating excess reserves (X = 0). The tiny slivers between the two lines contain combinations that give rise to either the “neither” (left side) or the “mixed” (right side) regime. Among other things, the figure implies that, even had the IOER rate been left at zero, banks would not have accumulated reserves so long as the effective federal funds rate itself remained at or above 6 basis points.
Finally, Figure IV.5, also taken from VanHoose and Dutkowsky, shows where the switching boundary lines shown in the previous figure stood, relative to the effective federal funds rate, between December 17, 2008 and December 17, 2015, when the IOER rate was fixed at 25 basis points. As can be seen, apart from several brief instances (during the first half of 2009 and the second half of 2010) in which it predicts a “mixed” equilibrium, VanHoose and Dutkowsky’s calibration exercise predicts that the Fed would have operated a floor system, and that it generally might have done so even had the IOER rate been considerably lower.
and the Effective Federal Funds Rate, December 2008 – December 2015.
(Source: Reproduced from VanHoose and Dutkowsky 2017, p. 13)

IV. e. The Floor System and the Friedman Rule

What became, in this transition from a corridor to a (leaky) floor system, of the Fed’s pre-crisis plan to use interest payments on reserves to eliminate the tax burden of reserve requirements, and to thereby discourage banks from expending valuable resources on sweep accounts and other devices for avoiding such requirements?

Most writings concerning the Fed’s floor system (e.g., Keister, Martin, and McAndrews 2015) seem to take for granted that the IOER rates the Fed has set in establishing and maintaining that system have been at least roughly consistent with Milton Friedman’s (1969) famous rule for achieving an “Optimum Quantity of Money,” according to which money balances ought to yield an average return competitive with that on other assets. Although Friedman, bearing in mind the practical impossibility of paying interest on currency, suggested that his rule might be satisfied by means of a deflation rate corresponding to the real rate of return on private financial assets, his rule can in principle also be satisfied without resorting to deflation, by having money assets, including bank reserves, bear interest.

Does it follow, then, that the Fed’s IOER policy is itself optimal? It doesn’t, for two reasons. First and most obviously, while there is a unique IOER rate consistent with satisfying the Friedman rule, any rate either equal to or above the Friedman-rule rate might serve to implement and maintain a floor system. In other words, although its move to a floor system meant that the Fed no longer taxed bank reserves, that move also made it possible for the Fed to subsidize bank reserve holding by paying a return on reserves exceeding that on other risk-free assets. We shall soon consider evidence that this has in fact been the case.10

---

10 It is sometimes said that, because reserves can be created “costlessly,” there is no risk of encouraging banks to hold too many of them. But while this claim is borne-out by some very simplistic models, it is never correct in practice. Even allowing that the nominal stock of bank reserves can be increased costlessly, the real stock can be increased only by reducing commercial banks’ share in total financial intermediation.
Second, a return on money below the Friedman-rule return is actually likely to be optimal under many common circumstances, such as when other taxes are also distortionary and when nominal prices are inflexible or “rigid” (Phelps 1973; Walsh 1984; Schmidt-Grohé and Uribe 2004). According to a recent estimate by Matthew Canzoneri, Robert Cumby, and Behzad Diba (2016), taking account of the costliness of nominal price adjustment alone, the optimal tax on excess reserves, instead of being zero, would be somewhere between 20 and 40 basis points in the steady-state. For much of the post-crisis period, this would have meant paying a negative interest rate on banks’ excess reserves, as several central banks eventually chose to do. Allowing for other externalities, or for the fact that other taxes tend to be distortionary, makes the optimal reserve tax even higher.

IV. f. A Dubious Advantage

Optimal tax or not, did the Fed’s new floor system at least save the day by allowing a rapidly-expanding Fed to maintain an above-zero fed funds rate? As I’ve observed elsewhere (Selgin 2017a), the logic underpinning the Fed’s plan was more than a little tortured. If there is reason to fear the zero lower bound, it’s because, once the fed funds rate reaches zero, banks, instead of seeking to exchange excess reserves for other assets, will become indifferent between those alternatives. “Banks,” Marvin Goodfriend (2002, p. 2) explains,

will never lend reserves to each other at negative (nominal) interest if reserve deposits are costless to store (carry) at the central bank. The zero bound on the

relative to the share handled by the central bank. Because commercial bank assets can include loans and securities that most central banks are (for good reasons) not allowed to possess, any increase in a central bank’s total intermediation share has real consequences; and those consequences are generally detrimental. Indeed, they can be so detrimental that an entire literature on “financial repression” is devoted to discussing them (Ito 2009). An ideal IOER policy must therefore steer a path between the Scylla of financial repression on the one hand and the Charybdis of suboptimal liquidity on the other; moreover, even such an ideal policy would neither justify nor make up for the inherent inefficiency of binding minimal reserve requirements—a point often overlooked in the literature. I offer evidence of the financially-repressive consequences of Fed’s floor system below.
nominal interbank rate is a consequence of the fact that a central bank stores bank reserves for free.

At the zero lower bound, ordinary Fed rate cuts may be practically or legally impossible. Those inclined to identify monetary easing with rate cuts see this as “the” problem. But that’s taking a superficial view of matters. The real problem is that, at the zero lower bound, assets that usually bear interest cease to be more attractive than money, placing the economy in a “liquidity trap.” In such a trap, St. Louis Fed economists Maria Arias and Yi Wen (2014) explain,

increases in money supply are fully absorbed by excess demand for money (liquidity); investors hoard the increased money instead of spending it because the opportunity cost of holding cash—the foregone earnings from interest—is zero when the nominal interest rate is zero.

In particular, banks will tend to hoard new reserves that come their way, instead of taking advantage of them to engage in more bank lending and deposit creation. As Congressman Alan Goldsborough famously put it in 1935, in attempting to induce more bank lending by adding to the stock of bank reserves, the Fed would find itself “pushing on a string.”

How, then, might a positive IOER rate help? To be sure, it can solve the “zero lower bound problem” superficially, by establishing a positive fed funds rate floor. But to what end? IOER would then render additions to the stock of bank reserves ineffective as a source of stimulus before the fed funds rate reached zero rather than once it did so. Yes, with the help of (positive) IOER, the Fed might set and achieve whatever positive rate target it liked; and it might do so regardless of how many reserves it created. But this “decoupling” of interest rates changes from changes in the scarcity of bank reserves,

---

11 Whether the U.S. economy was actually in a liquidity trap in the 1930s is controversial. For a compelling denial of the claim see Orphanides (2004).
applauded by Goodfriend (ibid.), Keister (2012), and Keister, Martin, and McAndrews (2008) as a feature of a floor system, is really a bug: the extra freedom it entails comes at a very great price, to wit: the Fed's inability to use its reserve-creating powers to promote additional bank lending and spending.  

When driving an automobile, one can get away with only so many combinations of steering-wheel movements on the one hand and the gas pedal pressure on the other. Wouldn't it be nice to be able to have complete freedom to step on the gas, and yet steer whichever way we like? Well, there's a solution: put the transmission in neutral! The hitch of course is that, while one can now steer any way one likes, and stomp on the gas all one likes, one cannot get very far doing either.

An above-market IOER rate can likewise allow the Fed to steer the fed funds rate any way it likes, while stepping on the reserve-creation pedal as hard as it likes, only by putting the usual monetary transmission mechanism in neutral. For the usual zero lower bound liquidity trap, it substitutes an above-zero liquidity trap in which monetary policy remains, despite appearances to the contrary, more or less equally impotent. The zero lower bound problem is thus avoided, but in a way that may still leave the economy in a depressed state, with little scope for monetary policy stimulus of the old-fashioned sort. It is as if (to offer one last simile), out of concern for would-be jumpers, the designers of a skyscraper decided to construct a broad concrete veranda around their building's second floor, to prevent them from ever hitting the ground!

Thanks to the Fed's switch to a floor system, and to its having kept that system in place by keeping the IOER rate above short-term market rates, the U.S. economy has been held in an above-zero liquidity trap since the trough of the Great Recession. Because

---

13 These remarks, once again, refer only to the use of a positive IOER rate to maintain an above-zero interest rate floor. A negative IOER rate can, in contrast, serve in principle to get around the zero lower bound problem by allowing a central bank to maintain a positive opportunity cost of reserve holding even when short-term market rates fall to zero. It is, to say the least, hardly possible that either negative or positive (but not zero!) IOER can serve equally well to get around the zero lower bound problem: if one theory of how IOER does this is correct, the other is, presumably, mistaken. I take up this point further in chapter IX.
of that, the Fed’s asset purchases have been much less effective than they might otherwise have been in promoting recovery and in keeping the inflation rate on target. Indeed, as Arias and Wen (2014) explain, because it involved large-scale purchases of long-term debt, and a consequent flattening of the yield curve, Quantitative Easing may actually have reinforced the floor-system based, above-zero liquidity trap.

***

Just how the Fed proposed to stimulate the economy with its ordinary monetary policy transmission mechanism stuck in neutral, as it were, was a challenging question Fed authorities would eventually have to answer. However, in early November 2008, stimulating the economy still wasn’t Fed officials’ main concern. Instead, their concern was to avoid stimulating the economy unintentionally. For that purpose, IOER, administered according to the requirements of a floor system, would serve the Fed’s needs well. In retrospect, it may be said to have served them all too well.
V. Digression: Is the Floor System Legal?

V. a. The Law’s Original Intent: A Below-Market IOER Rate

In switching from its circa-2006 plan to use interest payments on reserves for the modest purpose of ending the implicit taxation of bank reserves, to its new plan in which those payments would establish a floor system for monetary control, the Fed had to break, or at least bend, the law. For its new strategy marked a radical change, not just from what the authors of the 2006 legislation had envisioned, but from what that legislation allowed for in fact.

The pre-crisis opinion had been that interest on reserves should be used conservatively and cautiously. Most proposals called for interest to be paid on required reserves only, while all called for rates set low enough to avoid making reserves seem “more attractive relative to alternative short-term assets” (Weiner 1985, p. 30). Otherwise, the prevailing opinion held, interest on reserves, instead of simplifying monetary policy, could further complicate it (ibid).14

Such was clearly Federal Reserve Governor Laurence H. Meyer’s understanding when, in arguing the case for allowing the Fed pay interest on reserves before the House Banking Committee in 2000, he explained that

if the bill becomes law, the Federal Reserve would likely pay an interest rate on required reserve balances close to the rate on other risk-free money market instruments, such as repurchase agreements. This rate is usually a little less than the interest rate on federal funds transactions, which are uncollateralized overnight loans of reserves in the interbank market (Meyer 2000, p. 10).

What Governor Meyer considered an appropriate proxy for “the general level of short-term interest rates” in 2000 was presumably still appropriate in 2006. Since unsecured overnight rates, such as the federal funds rate and the London Interbank Overnight Rate (LIBOR), entail greater risk than overnight repos, to abide by the intent of

14 See also Laurent and Mote (1985).
the 2006 and 2008 laws, the Fed would have to keep the interest rate paid on reserve balances below these somewhat more risky overnight interbank lending rates. In this way, as one Fed official explained when the 2006 legislation was being considered, banks would have no reason “to significantly shift their financial resources to take advantage of this [the IOER] rate” (Eubanks 2002, p. 11). Instead, they would continue to keep only such reserve balances as they needed to meet their legal and clearing-balance requirements. The main difference reformers anticipated was that they would no longer bother using sweep accounts to avoid an implicit reserve tax.

The provisions of the 2006 legislation reflected these considerations. According to Title II of that measure, the Fed might pay interest on depository institutions’ reserve balances “at a rate or rates not to exceed the general level of short-term interest rates.”

V. b. Above the Law?

Except for a change in dates, the 2008 Financial Services Regulatory Relief Act left the details of the 2006 Act unaltered. Fed officials therefore found themselves in a quandary. As we’ve seen, they wanted to be able to resort to IOER three years ahead of schedule precisely for the purpose of making excess reserves “attractive relative to alternative short-term assets” (emphasis added). That meant setting the IOER rate above the going, but still positive, equilibrium fed funds rate. Indeed, given the “leakiness” of the Fed’s floor system, the IOER rate would have to be set considerably above the Fed’s target rate. That necessarily meant keeping the IOER rate above other, comparable market-based short-term interest rates. Yet by law, as we’ve seen, the Fed was only supposed to pay interest on bank reserve balances at a rate “not to exceed the general level of short-term interest rates.”

That the Fed’s IOER rates were well above comparable market rates can be seen in the next set of figures, the first of which (Figure V.1) compares the IOER rate to both the effective federal funds and the LIBOR rate:
Because the fed funds and LIBOR rates are rates for unsecured overnight loans, they include a small risk component, while the IOER rate is equivalent to a risk-free overnight rate. For that reason, and as Governor Meyers suggested in his previously-mentioned testimony, the rate implicit in overnight, Treasury-secured repurchase agreements might be a more appropriate market-rate benchmark. As Figure V.2 shows, that rate has also tended to fall below the IOER rate.

Finally, it’s instructive to compare the IOER rate to rates on Treasury bills of various maturities. The latter rates should, for obvious reasons, generally be above equivalent, risk-free overnight rates, according to the securities’ term to maturity. Yet, as the Figure V.3 shows, rates on both 4-week and 3-month T-bills have also been persistently, and often substantially, below the IOER rate. Indeed, from the spring of 2011

Figure V.2: IOR and OCR Treasury Repo Rates

Figure V.3: IOER, Overnight LIBOR, and Effective Federal Funds Rates
through mid-summer of 2015, even rates on 1-year Treasury bills remained below, and
generally well below, the IOER rate:

![Figure V.3: IOER and Treasury Bill Rates](image)

In short, it’s impossible to reconcile the Federal Reserve’s setting of its IOER rate
with any reasonable understanding of the law’s stipulation that it is “not to exceed the
general level of short-term interest rates.”

V. c. “One of these rates is not like the others...”

In an apparent after-the-fact attempt to legalize the Fed’s IOER rate settings, Fed
officials, in drafting the final rules implementing the 2008 statute, as announced in the
Federal Register on June 22, 2015, determined that for that purpose

“short-term interest rates” are rates on obligations with maturities of no more than
one year, such as the primary credit rate and rates on term federal funds, term
repurchase agreements, commercial paper, term Eurodollar deposits, and other
similar instruments (Regulation D: Reserve Requirements for Depository
Institutions 2015, p. 35567).

While most of the listed rates are unobjectionable, even if they fail to include
overnight obligations (which are, after all, closer equivalents to reserve balances than
term obligations are), the presence of the primary credit rate is a glaring anomaly, for
that’s the discount rate that the Fed charges sound banks for short term emergency loans.
As such it isn’t a market rate at all but one set administratively by the Fed’s Board of Governors. Moreover, since 2003 the Fed has always set its primary credit rate “above the usual level of short-term market interest rates” (Board of Governors 2017b). Since the Fed began paying interest on reserves it has also deliberately set its primary credit rate above the IOER rate.\textsuperscript{15} The Fed has thus found a way by which to claim, with an implicit appeal to Chevron deference, that its IOER rate settings have after all been consistent with the requirements of the 2006 law!\textsuperscript{16}

That the Fed should thumb its nose thus at the statute granting it the authority to pay interest on reserves would be regrettable enough if its doing so had only benign consequences. But that’s far from being the case. On the contrary: by bending the law to conform to its plan to make the accumulation of reserve balances more attractive to banks than other forms of investment, the Fed fundamentally altered the workings of the U.S. monetary system, with grave consequences for the U.S. economy.

\textsuperscript{15} Since the beginning of 2010 the Fed has maintained a fixed spread of 50 basis points between the IOER rate and the primary credit rate by adjusting both rates together.

\textsuperscript{16} “Chevron deference” is the controversial principle, put into effect by the Supreme Court’s 1984 decision in Chevron USA v. Natural Resources Defense Council, Inc., that courts should defer to government agencies’ own interpretations of statutes establishing new agency obligations and powers. In City of Arlington v. FCC (2013) the Court held, furthermore, that government agencies deserve deference even when it comes to interpreting statutes establishing the scope of their own authority!
VI. The Floor System and Interbank Lending

VI. a. Goodbye Fed Funds Market

As we’ve seen, when the Fed began paying interest on bank reserves, its immediate concern was to keep its emergency lending from causing the fed funds rate to drop below 1.5 percent—the target it set when it announced its IOER plan. To repeat Ben Bernanke’s words, “by setting the interest rate we paid on reserves high enough, we could prevent the federal funds rate from falling too low, no matter how much [emergency] lending we did (Bernanke, 2015).”

But interest on reserves could not discourage banks from placing newly-created reserves into the fed funds market without discouraging them from supplying any funds to that market: if a dollar of reserves that landed in a bank’s Fed account as a result of the Fed’s post-Lehman emergency lending earned more sitting in that account than it could earn if lent to another bank overnight, the same was true of a dollar of reserves held beforehand. Consequently, as Figure VI.1 shows, IOER served, not only to keep fresh reserves from lowering the fed funds rate, but to dramatically reduce the total volume of lending on the fed funds market: whereas financial institutions lent over $200 billion on the fed funds market during the last quarter of 2007, by the end of 2012 that figure has fallen to just $60 billion (Afonso, Entz, and LeSueur 2013). Although the figure has since returned to the neighborhood of $90-$100 billion, allowing for concurrent growth in overall nominal spending, the fed funds market remains relatively quiescent.

The outcome was, after all, fully predictable. Indeed, William Whitesell (2006, p. 1183), who was then Deputy Associate Director of the Board of Governors’ Division of Monetary Affairs, had predicted it not long before the crisis. Commenting then on the potential advantages of a floor system relative to a symmetric corridor system, he observed that, while the former, with its interest rate on reserves equal to the target

---

17 Merely increasing the quantity of reserves would not, on the other hand, have led to any decline in interbank lending so long as the Fed saw to it “that there is an opportunity cost to holding reserves, by remunerating them at a rate below the market rate” (Borio and Disyatat 2009, p. 18 n29).
interest rate, would have the advantage of insulating interest rates from the central bank’s reserve supply errors, the downside was that “trading in the overnight market might dry up,” with potentially “deleterious effects on market functioning.”

Once the new floor system was in place, banks and bank holding companies that were eligible for IOER almost completely stopped lending overnight funds. Only the Federal Home Loan banks and other GSEs continued to lend as much as ever, for the sake of securing a share of banks’ IOER earnings. The fed funds market thus ceased to function, as it had for decades, as banks’ preferred and most reliable source of last-minute liquidity, having instead been transformed to a substantial degree into a mere vehicle for bank-to-GSE interest-rate arbitrage.

Because banks that were once regular participants in the fed funds market no longer had reason to manage their liquidity by means of overnight lending and borrowing, they also lacked an important, former reason to scrutinize one another, and to thereby contain systemic risk, to wit: a desire to avoid incurring losses on unsecured loans (Rochet and Tirole 1996). As Craig Furfine (2001, p. 34) observes, when banks were inclined to “lend significant amounts of money to one another every day in the federal funds market,” as they did until the advent of interest on reserves, they had “an incentive to monitor their counterparties and to price these loans as a function of…the credit risk of the borrowing bank.”
Furfine’s investigation of rates charged in actual fed funds transactions shows, furthermore, that those rates really did reflect the varying credit risk of different institutions, with “banks with higher profitability, higher capital ratios, and fewer problem loans pay[ing] lower rates” than others (ibid.) Banks were able, in other words, to “distinguish credit risk among their peers and price loan contracts accordingly” (ibid., p. 54). Furfine notes as well that observed risk premiums were limited to banks that were not considered all that likely to default. For banks “with a significant perceived possibility of default,” the federal funds market “dries up” altogether (ibid.), in part because banks that might only secure funds at very high rates would rather not enter the market than signal their difficulties to other banks.

VI. b. IOER vs. Perceived Counterparty Risk

The fact that very risky borrowers are likely to be altogether excluded from the fed funds market may appear to lend credibility to the claim that the post-Lehman decline in fed funds lending was due, not to interest on excess reserves, but to a persistent post-Lehman increase in perceived counterparty risk. But perceived counterparty risk is no more capable of explaining the persistent decline in interbank lending than it is capable of explaining banks’ persistent accumulation of excess reserves, which I discuss at length in the next chapter. Although the TED spread—a popular measure of the perceived counterparty risk, equal to the difference between the interest rate on short-term interbank lending and the interest rate on Treasury securities—spiked at the time of Lehman’s failure, it began to decline soon afterwards when the Fed decided to come to AIG’s rescue, eventually falling to levels even lower than those that that prevailed before the crisis. Interbank lending, on the other hand, never recovered. The Fed’s decision to pay interest on excess reserves therefore appears to have been the fundamental cause of the enduring post-Lehman decline in such lending.\footnote{Bech et al. (2015) offer interesting insights concerning the combined effects on wholesale lending of the fears raised by Lehman’s failure and the Fed’s IOER announcement.}
Finally, the close relationship between the total volume of interbank lending and the opportunity cost of reserves holding, as measured by the difference between the interbank lending rates and the IOER rate, also supports the view that IOER drove the decline in interbank lending. Although the relationship is similar for all banks, it is clearest for foreign banks which, as we’ve seen, have been especially inclined to accumulate excess reserves. As Figure VI.2 shows, the precipitous decline in the opportunity cost of reserves, brought about by the Fed’s decision to pay interest on excess reserves, coincided with an equally precipitous, initial decline in foreign-bank interbank loans.

![Figure VI.2: Interbank Lending and the LIBOR-IOER Spread](image)

**VI. c. From Lender to Borrower of First Resort**

The collapse of interbank lending created a further motive, beyond the return on reserves itself, for banks to accumulate excess reserves, as banks that once routinely relied on overnight unsecured loans to meet their liquidity needs discovered that, owing to the substantial decline in the availability of fed funds, doing so was no longer prudent. Because that decline at first caught many banks by surprise, its immediate effect was a sharp spike, on October 7, 2008, in the fed funds rate, which rose to 2.97 percent, or almost twice the Fed’s target at the time. Banks adapted by raising their excess reserve holdings so as to have sufficient precautionary reserves to cover those reserve needs that they had previously met by borrowing federal funds.
As Gara Afonso, Anna Kovner, and Antoinette Schoar (2011, p. 1109) point out, until these changes came about, the fed funds market had long served as “the most immediate source of liquidity for regulated banks in the U.S.” Consequently any disruption of that market could “lead to inadequate allocation of capital and lack of risk sharing between banks.” In extreme cases, they add, it might “even trigger bank runs.”

In short, by paying IOER at above-market rates, and thereby establishing a floor system, the Fed, which is supposed to serve as a lender of last resort, became instead a *borrower* of first resort, destroying in the process banks’ traditional, first-resort source of short-term, emergency funds.
VII. The Floor System and Reserve Hoarding

Many observers still assume that the seemingly modest rate the Fed has paid on banks’ excess reserve balances, which is now 150 basis points, but was a mere 25 basis points from December 2008 until December 2015, has never been high enough to dramatically influence banks’ portfolio allocations, much less either the total supply or allocation of credit.

But as we’ve seen, these seemingly low IOER rates have not been low relative to comparable market rates. For that reason, their influence on banks’ behavior has been anything but modest. As Simon Potter (2015), a Federal Reserve Bank of New York Vice President, and head of its Market Group, explains,

The IOER rate is essentially the rate of return earned by a bank on a riskless overnight deposit held at the Fed, thus representing the opportunity cost to a bank of using its funds in an alternative manner, such as making a loan or purchasing a security. In principle, no bank would want to deploy its funds in a way that earned less than what can be earned from its balances maintained at the Fed.

Thanks to IOER, banks have refrained from acquiring assets bearing a net return below what they might earn simply by retaining Fed reserve balances. Some, indeed, have found it worthwhile to actively acquire Fed balances for the sake of arbitraging the spread between the return on such balances and private-market borrowing costs.

VII. a. The Accumulation and Distribution of Excess Reserves

Apart from its effect on interbank lending, the most notorious consequence of the Fed’s interest payments on excess reserves has been unprecedented growth in banks’ excess reserves balances, meaning the Fed balances they hold beyond those serving, together with banks’ holdings of vault cash, to meet their minimum legal reserve requirements.

In the two decades prior to October 2008, banks generally held between $1 and $2 billion in excess reserves, in part for the sake of avoiding shortfalls from their required
reserves, but mainly to avoid relatively costly clearing overdrafts. (The few exceptions consisted of short-lived spikes in excess reserves following crises, like that of September 11, 2001, when banks briefly held over $19 billion in excess reserves.) Banks’ minimum reserve requirements were, in contrast, largely met by their holdings of vault cash. Between them, minimum reserve requirements and banks’ demand for excess reserves for settlement purposes determined banks’ overall need for reserve balances, together with their desired ratio of such balances to their demand deposits. As Figure VII.1 shows, reserve balances normally amounted to between one-fifth and two-fifths of one percent of demand deposits only.

Figure VII.1: Demand Deposits, Reserve Balances, and Balances/Deposits, 1999-2008

As the next figure (VII.2) shows, after Lehman Brothers’ failure, banks’ excess reserve holdings began growing in lock-step with growth in the Fed’s balance sheet, starting with growth fueled by the Fed’s post-Lehman emergency lending, and continuing, after December 2008, with its several rounds of large-scale asset purchases (LSAPs). By August 2014 excess reserves, which had rarely surpassed $2 billion before the crisis, had risen to almost $2.7 trillion.
The sharp increase in perceived counterparty risk that immediately followed Lehman’s failure itself accounted, to be sure, for some increase in banks’ desire for excess reserves at that time. But as I noted in in the previous chapter, in discussing the collapse of interbank lending, the increase in perceived risk, as measured by the TED spread, was only temporary; and, as Figure VII.3 suggests, that temporary increase can’t account for banks’ willingness to continue accumulating excess reserves long after the panic had subsided.
That an above-market IOER rate alone could cause banks to cling to any new reserves that came their way should not have surprised anyone. As Finadium’s Jonathan Cooper (2012) observed in August 2012, although the Fed was only paying 25 basis points on reserves, that was enough to keep “lots of cash out of the securities market” since “You have to go out to 2 year notes before UST rates match the 25 bp that the Fed is paying.”

Still, banks didn’t all take part equally in the vast reserve buildup. Instead, as Figures VII.4 and VII.5 show, most of the new reserves ended up at the very largest U.S. banks or at U.S. branches of foreign banks. As of early 2015, the top 25 U.S. banks, by asset size, held more than half of all outstanding bank reserves, with the top three alone holding 21 percent of the total, while foreign bank branches accounted for most of the rest. The cash assets of small U.S. banks, in contrast, rose only modestly.

Figure VII.4: Excess Reserves, by Bank Asset Size, 2006-2015

---

10 For theories, see Dutkowsky and VanHoose (2017) and Ireland (2012). According to the latter’s DSGE model, in the absence of positive costs of managing large excess reserve holdings banks receiving interest on reserves at an above-market rate will wish to hold “an unboundedly large stock of reserves.” To avoid that outcome the IOER rate must be set slightly below the market rate (ibid., pp. 28–9). Bewley (1980) and Sargent and Wallace (1985) were among the first authors to draw attention to the problem of reserve demand indeterminacy in an IOER regime with a return on bank reserves equal to that on non-reserve assets.
That the very largest banks secured such a large share of the total accumulation of excess reserves is partly explained by the fact that those banks include some of the primary dealers that served as the Fed’s immediate counterparties in its asset purchases (Craig, Millington, and Zito 2014). Having thus had “first dibs” on new reserves the Fed created, primary dealer banks simply refrained from letting go of reserves they acquired. That practice was, of course, quite contrary to what primary dealers were normally expected to do, and to what they generally did do before the crisis, when the Fed was still relying on its traditional means of monetary control. Indeed, in the early stages of the subprime crisis, Fed officials worried that the collapse of ailing primary dealers would prevent them from serving as reliable conduits through which fresh reserves would make their way from the Fed to the rest of the banking system (e.g., Kohn 2009). Now, paradoxically, IOER was itself serving to close the same conduits, along with much of the rest of the interbank market, but was doing so deliberately as part of the Fed’s new monetary control strategy.

As for U.S. branches of foreign-owned banks, the vast majority of them aren’t subject to FDIC premium assessments. U.S. banks, on the other hand, are generally subject to those assessments. Moreover, since April 2011 FDIC premiums, which were previously assessed against banks’ U.S. deposits, have been assessed against their total assets, including their excess reserve balances, less tangible equity. For that reason, and

![Figure VII. 5: Excess Reserves, Foreign and Domestic Banks, 2006-2015](image)
also because many of their (mostly European) parent companies enjoy much lower net interest margins than U.S. banks, they’ve found it especially profitable to acquire fed funds for the sake of arbitraging the difference between the Fed’s IOER rate and lower private-market interest rates. In consequence these banks ended up playing a particularly important part in keeping growth in the total quantity of reserve balances from contributing to corresponding growth in overall bank lending.20

That very large U.S. banks and foreign bank branches have been especially inclined to hoard Fed reserve balances does not mean that small banks refrained altogether from doing so. As we’ve seen, by cutting off the flow of interbank funds, the Fed’s above-market IOER rate made it necessary for banks that still found excess reserves less remunerative than other assets to accumulate such reserves to protect themselves against the risk of occasional, large reserve outflows, such as go hand-in-hand with lending. As the late Ronald McKinnon (2011) observed in a Wall Street Journal Op-Ed,

Banks with good retail lending opportunities typically lend by opening credit lines to nonbank customers. But these credit lines are open-ended in the sense that the commercial borrower can choose when—and by how much—he will actually draw on his credit line. This creates uncertainty for the bank in not knowing what its future cash positions will be. An illiquid bank could be in trouble if its customers simultaneously decided to draw down their credit lines.

Ordinarily, McKinnon continued, banks can cover their unexpected reserve shortfalls by borrowing funds from other banks on the interbank market. However, if “large banks with surplus reserves become loath to part with them for a derisory yield,” while smaller ones “cannot easily bid for funds at an interest rate significantly above the prevailing interbank rate without inadvertently signaling that they might be in trouble,” interbank borrowing ceases to be an attractive alternative to maintaining higher excess reserve

20 According to Ayelen Banegas and Manjola Tase (2017, p. 14), as a result of the April 2011 change in FDIC fee assessments, U.S. banks’ net return on reserve balances initially declined by about 10 basis points. Since then, however, reduced FDIC assessments reflecting banks’ improved condition, have reversed the greater part of that decline.
cushions, even where the marginal return on reserves is less than that on loans. In fact, since late 2008, the “prevailing interbank rate” was irrelevant: to get other banks to lend to them on the federal funds market, smaller banks would have had to offer more than the going IOER rate, which was itself higher than short-term market rates. Needless to say, virtually none of them ever did.

VII. b. Excess Reserves and the Fed’s Balance Sheet

Some economists have insisted that, instead of depending on the Fed’s decision to pay interest on excess reserves, the post-Lehman accumulation of excess bank reserves was an inevitable consequences of the Fed’s asset purchases. In an influential Liberty Street post, for example, Gaetona Antinolfi and Todd Keister (2012) criticized Alan Blinder (2012) and others for claiming that lowering the IOER rate would encourage banks to lend more and thereby reduce their excess reserve balances:

Because lowering the interest rate paid on reserves wouldn’t change the quantity of assets held by the Fed, it must not change the total size of the monetary base either. Moreover, lowering this interest rate to zero (or even slightly below zero) is unlikely to induce banks, firms, or households to start holding large quantities of currency. It follows, therefore, that lowering the interest rate paid on excess reserves will not have any meaningful effect on the quantity of balances banks hold on deposit at the Fed. ... In fact, the total quantity of reserve balances held by banks conveys no information about their lending activities—it simply reflects the Federal Reserve’s decisions on how many assets to acquire (Keister and Antinolfi 2012).

It’s of course true, as any money and banking textbook will affirm, that banks cannot alter the total quantity of reserve balances simply by trading them for other assets, as doing so only transfer the balances to other banks. But the question isn’t whether a lower IOER rate would reduce total reserves. It’s whether it would result in a lower the quantity of excess reserves. The answer to that question is “yes,” because, as the same textbooks also explain, as banks trade unwanted reserves for other assets, they also
contribute to the growth of total banking system deposits; the fact that unwanted reserves get passed on like so many hot potatoes only makes deposits grow that much more rapidly. The growth of total deposits serves in turn to convert former excess reserves into required reserves, where “required” means required either to meet minimum legal requirements or for banks’ clearing needs.

That, at least, is what always happened before the Fed began encouraging banks to cling to excess reserves. For example, as Figure VII.6 shows, until October 2008 banks routinely disposed of unwanted excess reserves in the manner just described, thereby keeping system excess reserves at trivial levels, and doing so despite additions to the total supply of bank reserves that were, by pre-2008 standards at least, far from trivial. For example, as Donald Dutkowsky and David VanHoose (2017, p. 1) observe, “from January 1997 to September 2008 the monetary base nearly doubled, from $463 billion to $910 billion. Yet the share of bank assets held as excess reserves remained near zero throughout.”

![Figure VII.6: Total and Excess Reserve Balances, 1984-2008](image)

It follows that, when banks hold a large quantity of excess reserves, that fact actually conveys very significant “information about their lending activities.” Specifically, it tells us that something has caused them to refrain from engaging in such activities precisely to the extent that they choose to accumulate reserves instead.
In reply to these observations Professor Keister has suggested that, floor system or no floor system, the unprecedented scale of the Fed’s post-Lehman balance sheet growth would have rendered the traditional means by which banks disposed of unwanted excess reserves inoperable, because banks couldn’t possibly achieve the expansion in their total assets and deposit liabilities required to convert so vast an increase in total reserves into an equally vast increase in required reserves. But this counter-argument is also contradicted by relevant historical evidence, consisting of instances of hyperinflation in which central banks expanded their balance sheets on a scale much larger still than that seen in the U.S. since 2008. During the notorious Weimar hyperinflation, for example, the (proportional) growth in German bank reserves far exceeded that witnessed in the U.S. since Lehman’s bankruptcy. Yet, according to Frank D. Graham (1930, p. 68), Germany’s banks, far from accumulating excess reserves, increased their lending more than proportionately. “It would appear,” Graham writes, “that the commercial banks extended loans throughout the period of post-war inflation considerably in excess of a proportionate relationship with the increase in the monetary base. ... The increase in deposits issuing from loans was especially marked in 1922 and till stabilization in 1923.”

It doesn’t follow, of course, that, had it not been for interest on excess reserves, the Fed’s post-Lehman asset purchases would have led to hyperinflation. Instead, Fed officials would not have dared to purchase assets on such a large scale; alternatively, had they made the attempt, they would have ended the purchases once confronted with evidence that the inflation rate was in danger of exceeding its target. As it was, by relying on IOER to discourage banks from dispensing with excess reserves, the Fed ended up falling short of, instead of surpassing, its inflation target. That outcome came as a surprise to those accustomed to the workings of the Fed’s traditional monetary control framework. But in the context of its new floor framework, any tendency for the Fed’s asset purchases to raise prices would itself have been surprising.

---

21 In personal correspondence.
VII. c. Excess Reserves and the Liquidity Coverage Ratio

Others may wonder whether, rather than being due to excess reserves’ high yield relative to other assets, banks’ continuing willingness to hold substantial quantities of such reserves is a result of their need to satisfy Basel’s Liquidity Coverage Ratio (LCR) requirements. The Basel requirements, which were first applied to U.S. banks at the beginning of 2015, call for bank holding companies having at least $250 billion in assets, or ones with $10 billion or more in foreign exposure, to maintain a level of “High Quality Liquid Assets” (HQLAs) against their “non operating” deposits equal to 100 percent of their 30 day liquidity outflows.\(^{22}\)

Although banks’ required reserves do not count as HQLAs for the purpose of satisfying LCR requirements, their excess reserves do. Consequently it’s possible that, with the Basel LCR requirements now in effect, even if reducing the IOER rate would have made a difference in the past, it would no longer lead to any substantial reduction in banks’ demand for excess reserves. Credit Swiss’s Zoltan Pozsar (2016, pp. 2-3), for one, appears to take this view. “Contrary to conventional wisdom,” he says, “there are no excess reserves—not one penny”:

Labelling the trillions of reserves created as a byproduct of QE as “excess” was appropriate only until the Liquidity Coverage Ratio (LCR) went live, but not after. ... Before the LCR, excess reserves were indeed excess: every penny was in excess of the amount of reserves required by the Federal Reserve’s Regulation D. Under the LCR, all excess reserves became required: not to comply with Regulation D, but with the LCR.... It is helpful to think about the LCR as a global reserve requirement regime.... Excess reserves are not sloshing but rather sitting at the Fed. They sit passive and inert because banks must hold these reserves as HQLA to meet LCR requirements.

\(^{22}\) Operating deposits include retail deposits and wholesale deposits held for clearing, custody or cash management purchases. All other unsecured wholesale deposits are considered non-operating.
Pozsar goes on to conclude that the new requirements imply “the need for a big Fed balance sheet for a long time to come” (ibid, p. 2), if not forever (ibid, p. 9). “There is no turning back,” he adds, “to the old days where reserves were scarce. The LCR does not allow that... Big is necessary. It is the future. Get over it...” (ibid, p. 12).

But is it true that the LCR “does not allow” banks to get away with modest quantities of excess reserves? It isn’t. While such reserves qualify as HQLAs, and, specifically, as the highest quality “Tier 1” HQLAs, so do U.S. Treasury securities, Ginnie Mae MBS, non-GSE agency debt, and, since October 2014, deposits at the Fed’s Term Deposit Facility. Furthermore, up to 40 percent of banks’ HQLAs may consist of “Tier 2” assets, which include other GSE securities, certain corporate bonds, and qualifying common stock. In short, banks might, in principle, meet their LCR requirements without holding any excess reserves at all. It turns out that in claiming that banks are now “required” to hold immense quantities of excess reserves to meet their LCR requirements, what Pozsar really means is that they prefer to meet the requirements that way given what those reserves have been yielding relative to other HQLAs! Cumberland Advisors’ David Kotok (2016), in a post praising and elaborating upon Pozsar’s arguments, makes this perfectly clear:

The LCR requirement is met by the election of the commercial bank. Each bank, pricing its available alternatives, determines how to comply. ...[T]he current worldwide pricing of alternatives favors the use of reserve deposits at the Fed. That explains why about half of the excess reserves at the Fed are placed there by U.S. subsidiaries of foreign banks. Those foreign-owned deposits meet LCR. At the same time those banks are earning 50 bps paid in U.S. dollars instead of paying 40 bps in euro. That 90-bps spread is serious money and may be changing agents’ behavior.

---

23 This was written on June 9, 2016.
Changing behavior indeed. A recent study by Jane Ihrig and several coauthors, using standard portfolio theory and data for the 2012-2016 period, finds that, even allowing that Treasuries are somewhat riskier than reserves, large U.S. banks “should not want to hold any amount of the risk-free asset—reserve balances—to satisfy HQLA” (Ihrig et al. 2017, p. 15). That finding is, moreover, robust for all values (1 being the lowest and 10 the highest) of their model’s risk-intolerance parameter. A follow-up analysis, using the longer 2001-2016 sample period to allow for greater volatility of asset returns, still found Treasuries and GSE MBS dominating reserves for modest (<5) values of the risk-intolerance parameter, and that despite the counterfactual assumption of an IOER rate set equal to the effective fed funds rate prior to 2008. Finally, for smaller values of risk intolerance, even a small change in the IOER rate might result in a substantial change in banks’ demand for excess reserves. If, for example, the risk-intolerance parameter a has a value of 3, a 25 basis point reduction in the IOER rate would, all else equal, reduce excess reserves’ share of HQLAs “by about 10 percentage points,” while increasing that of Treasury securities by the same amount (ibid, p. 17). Overall, these findings suggest that IOER has contributed substantially to banks’ use of excess reserve balances to meet their LCR requirements.\(^{24}\)

\textit{VII. d. Reserve Demand and Opportunity Cost}

Final proof, should it be needed, of the bearing of IOER on banks’ willingness to accumulate excess reserves comes from consideration of how that willingness has varied with changes in the relationship between the IOER rate and corresponding market rates. If banks’ demand for excess reserves is driven by the yield on such reserves compared to that on other assets, then the ratio of excess reserves to total bank assets should vary with

\(^{24}\) According to Ihrig et al. (2017, p. 12), at the end of 2014, when the requirements were about to be implemented, the excess reserves of banks subject to them were sufficient to satisfy 45 percent of the requirements. Since then that share has fallen to about 36 percent. Excess reserve balances’ share of total HQLAs also varies considerably from bank to bank. Of eight of the largest bank holding companies, reserves’ HQLA share was 20 percent or less, whereas for some it was above 50 percent (ibid., pp. 18-19). Both the large holdings and the variations almost certainly reflect the degree to which reserves used to satisfy LCR requirements have been borrowed overnight from GSEs, as part of the fed funds arbitrage discussed previously.
the difference between the IOER rate and comparable short term market rates, such as the overnight LIBOR rate. A recent study by David Beckworth (2018) shows that this has indeed been the case. As Figure VII.7 shows, for U.S. banks as a whole, the two values are closely correlated, with \( R^2 = 77.24 \). As might be expected, the relationship for foreign banks, shown in Figure VII.8, is even stronger, with \( R^2 = 82.04 \) percent.

A multiple regression analysis, also performed by Beckworth, with reserves’ share of banks’ total assets as its dependent variable, and the IOER-LIBOR spread as well as several control variables as its independent variables, also showed the spread to be highly significant, both statistically (at the 1% significance levels) and economically, with spread coefficients ranging from .094 to small domestic banks to .554 for larger ones to just shy of 2 for foreign bank branches. These findings once again agree with our understanding that larger U.S. banks stood to profit more by acquiring reserves than small ones could, and that foreign bank branches profited most of all.
VIII. The Floor System and Retail Lending

VIII. a. Lending Before and Since the Crisis

Between the week just before the Fed began paying interest on bank reserves, when it reached its pre-crisis peak, and the third week of March 2009, when it reached its post-crisis nadir, overall U.S. commercial bank lending fell from over $7.25 trillion to about $6.5 trillion—a decline of $1.25 trillion. As Figure VIII.1 shows, although reduced real estate lending accounted for the greatest part of this decline, other kinds of lending, including business lending, also fell sharply.

![Figure VIII.1: Commercial Bank Lending, 1996-2017](image)

Although lending has since recovered to a considerable extent, at least relative to its pre-subprime boom trend, that recovery was painfully slow. It also masks an enduring and substantial post-crisis decline in the ratio of overall bank lending (“loans and leases”) to total bank deposits. As Figure VIII.2 shows, whereas total bank lending tended to match total bank deposits in the years leading to the crisis, since then, and specifically since IOER was introduced, it has declined to about 80 percent of deposits. Over that same period, bank reserves, as a percentage of total bank deposits, have increased from trivial levels to roughly 20 percent of bank deposits. In short, as a matter of simple balance-sheet arithmetic, the rise in banks' holdings of (mainly) excess reserves has gone hand-in-hand with a corresponding decline in bank lending.
VIII. b. The Direct Influence of IOER on Bank Lending

But does this correspondence mean that IOER was actually responsible for the decline in retail bank lending as a share of bank deposits? Many insist that IOER rates have been too low, compared to the rates on commercial bank loans, to have had more than a minor influence on bank lending. For example, Ben Bernanke and Donald Kohn (2016) observe that, during the long interval when the IOER rate stood at 25 basis points, “the only potential loans that would have been affected by the Fed’s payment of interest [on reserves] are those with risk-adjusted short-term returns between precisely zero and one-quarter percent.”

That view is, however, mistaken, on both empirical and theoretical grounds. First of all, as we’ve seen, the growth in banks’ excess reserve holdings was not an inevitable response to growth in the Fed’s balance sheet: banks are always materially capable of reducing their excess reserve holdings, collectively as well as individually, either by making loans or by buying securities. It follows that the existence of substantial excess reserve balances is ipso-facto proof that the banks that acquired those reserves considered them more desirable than any other assets they might have acquired.

Standard microeconomic theory suggests, furthermore, that in equilibrium all of a banks’ various assets should yield the same marginal return. Consequently, in theory at least, for any bank that holds excess reserves, the return on the marginal loan should be equal to the return on reserves. Allowing that the demand for loans is declining in the
loan rate of interest, this in turn implies that, if the return on reserves goes up, total bank lending must decline enough to once again equate the marginal return on loans with the (fixed) return on reserves. Although a decline in wholesale lending, and interbank lending especially, may be the first and most obvious consequence of an increase in bank reserves’ relative yield, the eventual consequences will also include a reduction in retail.

But can this orthodox theory account for a *substantial* decline in lending as a share of bank deposits? It can, provided one understands, first of all, that not all banks enjoy equally high returns on lending. That fact is at least roughly reflected in different banks’ net interest margins: the difference between the interest they earn and the interest they pay on bank deposits, expressed as a percentage of bank assets. Because bank deposit rates have been extremely low since the crisis, especially at larger banks, banks’ net interest margins supply a rough indication of their gross interest returns. As figure VIII.3 shows, those margins have long tended to be lower for the very largest U.S. banks (proxied here by New York banks), and lower still for Euro area banks (branches of which hold substantial quantities of excess reserves), than they have been for U.S. commercial banks as a whole; and the gap between small and large bank NIMs increased further since the crisis. Between August 2008 and mid-2015, NIMs of large (> $50 billion) U.S. banks fell 70 basis points, from about 330 basis points to about 280 basis points, mostly owing to declining yields on loans, while those of smaller banks declined only slightly (Covas, Razende, and Vojtech 2015). The NIMs of the very largest U.S. banks have fallen even lower. As we’ve seen, those banks, together with U.S. branches of foreign banks (which generally have still lower NIMs) have held the lion’s share of outstanding excess reserves.
Even 150 basis points is many times 25 basis points. But that’s still not the right comparison, because there are substantial non-interest expenses, including variable “balance sheet costs,” involved in making loans, whereas the sole balance sheet cost of holding Fed balances consists of FDIC premiums assessed against a bank’s total assets—and even that cost does not apply to most foreign bank branches. ECB area bank operating expenses, for example, are equal to about 60 percent of their interest income. Because borrowers sometimes default, and banks must make allowances for such defaults, loan loss provisions further reduce the net return on bank loans (Noizet 2016). As Figure VIII.4 shows, those provisions reached a peak of 3.7 percent of total bank assets at the beginning of 2010, from which they’ve gradually fallen to their present level of 1.29 percent. Taking such losses as well as other costs of lending into account, it’s no longer so difficult to understand how a modest IOER rate might have made holding excess reserves seem more lucrative than granting a loan at a considerably higher non-risk-adjusted rate.
Nor is that all. Banks’ net interest margins are a measure of the return on their entire loan portfolios. But because the demand schedule for bank loans is downward-sloping, the return on a bank’s marginal loan is necessarily lower than that on its loan portfolio as a whole; and it’s this marginal return that’s supposed, in equilibrium, to be no higher than the marginal return on other bank assets, including excess reserves. Consequently, the mere existence of a positive difference between a banks’ net interest margin and the IOER rate, even after allowing for the noninterest cost of loans, is perfectly consistent with the theory that banks have found it more profitable to accumulate excess reserves than to part with those reserves by lending more.

Figure VIII.5 illustrates the last point. In it, the blue line represents the downward-sloping marginal revenue schedule for loans confronting the banking system, while the horizontal grey line represents the IOER rate, here assumed to be 100 basis points. For simplicity, I ignore banks’ noninterest expenses altogether, while assuming that the Fed adjusts the total stock of reserves so as to keep total bank deposits constant.

Figure VIII.5: Bank Lending and IOER

In that case, assuming that they have $10 trillion in deposits at their disposal, the banks will collectively lend $8 trillion, while maintaining $2 trillion in excess reserves. But although the net return on the marginal loan is the same as the IOER rate, the banking system net interest margin, represented by the figure’s orange line, will necessarily be higher than the IOER rate. Reducing the IOER rate to zero, on the other hand,
encourages banks to lend 100 percent of their deposits, instead of holding any excess reserves.\footnote{Alternatively, one can treat the horizontal axis in the diagram as representing real rather than nominal bank deposits, where changes in the IOER rate lead to changes in the deposit multiplier and therefore to proportional changes in both nominal bank deposits and the price level. A simple, formal model that yields results consistent with the diagram can be found in Appendix I. Using a more elaborate model, Martin, McAndrews, and Skeie (2013) reach similar conclusions. In particular, they find that “the key determinant of bank lending is the difference between the return on [bank] loans and the opportunity cost of making a loan,” and that “banks lend up to the point where the marginal return on loans equals the return on holding reserves.” They also show that, once this point has been reached, further additions to the supply of bank reserves have no effect on bank lending; indeed, if banks’ balance-sheet costs are high enough, an increase in reserves can even lead to a \textit{decline} in bank lending. The last point has obvious implications for the likely effectiveness of the Fed’s Large-Scale Asset purchase. Andolfatto (2015), using yet another model, also reaches quite similar conclusions, as do VanHoose and Dutkowsky (2017), whose model I discussed previously.}

\textit{VIII. c. Excess Reserves and Bank Lending in Japan}

Some may doubt that IOER accounts for U.S. banks’ exceptional demand for excess reserves, and the associated decline in bank lending, because similar developments have been observed elsewhere where banks’ reserve balances bore no interest. Of these cases, Japan’s is perhaps the most notorious. As Kazua Ogawa (2007, p. 241) observes, “Japanese banks have chronically held excess reserves since the late 1990s,” with excess reserves tending to rise \textit{pari passu} with the central bank’s additions to the total reserve stock, just as happened in the U.S. after October 2008.

However, Ogawa also observes that Japan is no exception to the rule that “reserve supply does not necessarily automatically create a demand for reserves,” and that Japan’s banks, no less than U.S. banks, “have their own motives for excess reserves.” The motives were, moreover, more-or-less the same in both cases.

U.S. banks, as we’ve seen, accumulated excess reserves because the positive return on those reserves was greater than the still-positive return on wholesale as well as some retail loans. Japanese banks, in contrast, began hoarding reserves long before the Bank of Japan began paying interest on reserves a month after the Fed’s did so, in November 2008.
But as the U.S. case itself demonstrates, what matters isn’t the absolute IOER rate, but how that compares to rates on alternative uses of bank funds. In Japan before November 2008, although the IOER rate was zero, the overnight uncollateralized call rate—Japan’s equivalent to the fed funds rate—had itself fallen to zero, making reserves and call loans very close substitutes despite the fact that reserves bore no interest. Furthermore, as Figure VIII.6 shows, the Net Interest Margin of Japanese banks as a whole has been less than half—and often less than a third—that of larger (> $15 billion) U.S. banks. Finally, the fact that Japanese depositors became increasingly leery of bank failures in the 90s finally tipped the scale in favor of reserves, as Japanese banks gained a further incentive to bolster their precautionary balances.

Figure VIII.6: Net Interest Margins of Japanese and Large (> $15B) U.S. Banks

As can be seen in the next pair of figures, reproduced from Bowman, Gagnon, and Leahy (2010, p. 33), during the days when the Bank of Japan paid no interest on banks’ reserve balances, Japanese banks accumulated excess reserves only after March 2001, when the Bank of Japan, in initiating its Quantitative Easing Program, allowed the call rate itself to fall to zero.\(^{26}\) When the BOJ ended that program five years later, while also increasing its lending rate, the call rate again rose above zero, causing Japan’s banks to reduce their excess reserve balances. Finally, in November 2008, by beginning another round of Quantitative Easing, and reducing its lending rate to 30 basis points, the Bank of

\(^{26}\) According to Bowman et al. (2011, p. 6), despite zero rates on BOJ deposits, banks preferred accumulating balances there to increasing their interbank deposits, which yielded small but still positive returns at the time, in part because BIS regulations considered BOJ deposits to be riskless, whereas those regulations required that 20 percent of interbank deposits be counted as risky assets against which capital had to be held.
Japan brought the call rate back down 10 basis points, while simultaneously beginning to pay banks 10 basis points on their reserve balances. It was because of that change that Japanese banks once again began accumulating excess reserves.27

Figure VIII. 7 and VIII.8: Bank of Japan Policy Rates, and Japanese Banks’ Excess Reserves, 1999–2009
(Source: Bowman, Gagnon, and Leahy 2010, p. 32)

In short, like the Fed after October 2008, the Bank of Japan saw to it, intentionally or not, that Japanese banks’ excess reserve balances rose and fell in lockstep with changes in the size of its balance sheet, which they would not have done had it maintained a positive spread between the call rate and the rate it paid on excess reserves. According to Ogawa’s estimates, had Japan’s call rate been 25 basis points rather than zero after 2000, even with no improvement in Japanese banks’ perceived financial health, banks’ subsequent demand for excess reserves might have been reduced by as much as 70 percent.

27 For the purpose of paying interest on banks’ reserve balances, the Bank of Japan established its Complementary Deposit Facility. Although that facility was originally supposed to expire on March 16, 2009, it has since been made permanent. Interestingly, since banks can only maintain excess reserves at the facility, the BOJ paid interest on excess reserves only, and not on banks’ required reserves. Japan’s IOER rate remained positive until January 2016, when the Bank of Japan introduced a “three tier” arrangement for Japanese banks’ account balances with it, in which one tier pays a positive, one a zero, and one a negative interest rate.
Thanks to the Bank of Japan’s strategy, and in agreement with our own understanding that the influence of IOER on bank lending will be greatest where bank net interest margins are lowest, Japan’s Quantitative Easing programs, instead of resulting in more lending by Japanese banks, had just the opposite effect, as can be seen in Figure VIII.9:

While it doesn’t contradict the claim that IOER can be a crucial determinant of banks’ willingness to accumulate excess reserves, Japan’s experience does cast doubt on the suggestion that a U.S. IOER rate of zero would have sufficed after 2008 to have kept banks there from hoarding excess reserves. Whether it would have depends on whether other U.S. short-term rates, and the effective fed funds rate in particular, would have remained above zero. If not, nothing short of a negative IOER rate would have served to preserve a positive opportunity cost of reserve holding. Even so, a zero IOER rate would have supplied less of an inducement for reserve hoarding than a positive one. More importantly, as we shall see, by 2010 Fed officials themselves were convinced that, had they returned the U.S. IOER rate to zero, the effective fed funds rate, despite falling further, would nevertheless have remained positive.

VIII. d. Bank Capital as a Constraint on Bank Lending

To insist that the Fed’s floor system contributed to the post-Lehman decline in bank lending, and especially to the decline in lending as a share of total bank deposits, isn’t to say that other developments played no part in that decline. Most obviously, a
decline in overall loan demand was part of the story. But to suggest that it was this decline rather than IOER that mattered, as many in the banking industry seem inclined to do, is to erect a false dichotomy: if banks reduced their loans while increasing their reserves, they did so, not simply because lending became less lucrative, but because it became so relative to the alternative of reserve hoarding. Had it not been for IOER, banks would have been far less inclined to prefer reserves to low-yielding loans. IOER and reduced loan demand thus worked together, like the blades of a scissor, to discourage banks from lending.

A shortage of bank capital might, on the other hand, have prevented banks from increasing their loans despite the presence of both abundance of excess reserves and favorable lending opportunities. As Huberto Ennis and Alexander Wolman (2011) explain,

As a readily available source of funding, high levels of reserves provide flexibility to a bank that is looking to expand its loan portfolio. However, loans (and risky securities) are associated with higher capital requirements than reserves. A bank that is holding reserves but is facing a binding capital constraint is thus unlikely to engage in a sudden expansion of lending. As with deposits, raising capital quickly can be costly. For this reason, even a bank that holds a high level of excess reserves may not be able to take advantage of new lending (or investment) opportunities (p. 276).

However, in their own study of this possibility, Ennis and Wolman find that, while many banks were indeed capital constrained during the Fed’s “first wave of reserve increases,” by the last quarter of 2009 bank capital had recovered to the point where, of $510 billion in reserves held by the biggest 100 banks, $485 billion were loanable. Barajas et al. (2010, p. 9) reach the same conclusion. By the end of 2011, finally, almost all of the reserves held by the same banks were loanable given existing capital requirements. In separate study also looking at larger banks and BHCs, Jose Berrospide and Rochelle Edge (2010) likewise found that changes in BHCs’ capital ratios had only modest effects on loan
growth. Instead of worrying about capital, banks and BHCs seemed more concerned about things like loan demand and risk. (Alas, Berrospide and Edge did not consider the possible influence of IOER.)

Nor does capital seem to have significantly constrained lending at the opposite end of the banking spectrum, where banks must usually rely on retained earnings to build capital. According to Jim Wilkinson and Jon Christensson (2011, pp. 43 and 46), who investigate lending by community banks in the Tenth Federal Reserve District between the start of 2001 and the end of 2009, programs established during the crisis for the purpose of placing funds into those banks’ capital accounts did so little to boost that lending that it would have been “more effective for policymakers to give money directly to small businesses in the form of grants or loans.”
IX. The Floor System and Monetary Policy

IX. a. IOER and Tight Money in 2008–9

Having considered the bearing of the Fed’s switch to a floor system on various sorts of bank lending, we’re now equipped to consider how it influenced the course of the subprime recession and subsequent recovery. In brief, the Fed’s decision to implement a floor system, and to maintain a correspondingly high IOER rate, contributed to both the recession itself and the slow pace of the subsequent recovery by serving as the instrument by which the Fed—whether wittingly or not—kept money too tight.

Although there were clear signs of trouble in the subprime mortgage market starting in early 2007, the recession to which those troubles eventually led didn’t officially begin until December 2007. As is true by definition of any officially-designated recession, that one was heralded by a substantial decline in various measures of overall real economic activity, and particularly in the growth rate of real GDP, that had been going on for several months.

As is typically, though not necessarily, the case, the recession also involved a similar, but even sharper, decline in nominal GDP, or total spending on goods and services. As can be seen in Figure IX.1, from a peak rate of over 7 percent during the boom, nominal GDP growth declined gradually to about 4.75 percent in the third quarter of 2007. It then fell precipitously, reaching a low just shy of minus 3.2 percent by the second quarter of 2009. And although the growth rate of spending recovered considerably over the next year, since mid-2010 it has never again reached 5 percent, and has often been less than 3 percent. In short, spending never made up the ground it lost during the recession’s first year.
While the connection between reduced spending and recession isn’t inevitable, it’s a strong one, for reasons that aren’t difficult to grasp. For in order not to be accompanied by some decline in real GDP, a decline in nominal GDP would have to be matched by a proportional decline in prices, as measured by the GDP deflator. To the extent that it isn’t, because prices are “rigid” or “sticky” or for any other reason, real GDP must also decline. In practice, a sharp and persistent decline in overall spending is bound to bring a recession.

The volume of spending itself depends on the quantity of money, however one chooses to measure it, and its velocity, which can be understood as an inverse measure of the public’s demand for money balances, expressed as a share of their total earnings. As Figure IX.2 shows, although the velocity of M2 was growing at the beginning of 2006, by 2007 it was declining. That decline became increasingly rapid, and especially so after Lehman Brothers failed, so that by mid 2009 M2 velocity was more than 11 percent below its level a year before. Although the quantity of M2 tended to increase as its velocity declined, the increase fell persistently and increasingly short of what was needed to maintain a steady growth rate of spending, let alone what it would have taken to restore spending to its original trend path. Instead, that growth rate fell steadily until, during the last quarter of 2008, it became negative.
In light of these statistics, it’s clear in retrospect that monetary policy became increasingly tight during 2007 and early 2008, and that this overtightening became pronounced during the last quarter of 2008 and the first quarters of 2009. Taking a 5 percent spending growth rate to represent the long-run trend, it’s equally clear that money remained too tight over the next several years to restore that spending growth rate, let alone make up for the fallen level of spending relative to where it would have been had the growth rate of spending never fallen below 5 percent.

The especially severe overtightening that followed Lehman’s failure reflected the FOMC’s desire to maintain the 2 percent fed funds rate target then still in effect. That target was, according to the committee’s reckoning, consistent with meeting the Fed’s inflation target, whereas anything lower risked surpassing that target. Finding that its emergency lending was undermining that chosen target, the Fed responded, as we’ve seen, by implementing IOER as a means for preventing any further “leakage” of its emergency credits into the fed funds market. IOER thus became the chief instrument by which the Fed aggravated, however inadvertently, the collapse in nominal spending that was already in progress, making the recession that much more severe. Commenting on
the Fed’s action not long afterward, blogger David Beckworth (2008) went so far as to compare the Fed’s mistake to the one it made in 1936–1937.28

Some years later another blogger, Scott Sumner (2017), having the advantage of hindsight, reached a verdict that was hardly less damning. “The decision to adopt IOR,” he writes (meaning, presumably, interest on excess reserves), “helped to prevent the Fed from achieving its policy goals, by making the Great Recession more severe than otherwise.” He continues,

The world would be a better place today if the Fed had never instituted its policy of IOR in 2008. I really don’t see how anyone can seriously dispute this claim. If you want to dispute the claim, what specific way did IOR make the world a better place? When the policy was adopted in 2008, the New York Fed explained it to the public as a contractionary policy. Can anyone seriously argue that the world would be worse off if monetary policy had been less contractionary in 2008–12? Why?

Fed officials were aware of the economy’s deteriorating state as they prepared to begin paying banks to hold reserves: it was that deterioration, after all, that convinced them to finally reduce the federal funds rate target from 2 percent to 1.5 percent (see Figure IX.3). Yet the Fed still went ahead, the very next day, with its IOER plan. The Fed chose, in other words, to ease monetary policy symbolically, while taking steps to prevent the reserves it was creating from actually contributing to a further lowering of the effective funds rate. The FOMC’s next and final rate cut under what still appeared to be, but was in fact no longer, its traditional monetary control regime, from 1.5 percent to 1 percent, was likewise largely symbolic, for by then the fed funds market, considered as a market for interbank lending, had already ceased to function.

28 In an action widely believed to have contributed to the “Roosevelt Recession” of 1937-38, Fed officials, fearing that a revival of bank lending would sponsor inflation otherwise, doubled banks’ minimum reserve requirements between August 1936 and May 1937 in order to transform banks’ then-substantial excess reserves into required reserve balances.
Thus far, at least, the Fed’s experiment was proceeding according to plan. For despite the economy’s ongoing decline, that plan called not for loosening monetary policy but for avoiding further loosening, along with the stimulus such loosening might provide, by preventing growth in the Fed’s balance sheet from encouraging additional bank lending. Before the end of November 2008, however, the Fed had concluded that the economy needed to be stimulated after all. The trouble was that, with the new floor system in place, achieving a monetary stimulus was only barely possible in theory, and lamentably difficult in practice.

IX. b. The Floor System and “Quantitative Easing”

The problem of course was that, so long as the IOER rate remained high relative to other short-term rates, as the Fed’s floor system required, the Fed’s asset purchases, no matter how large, would tend to lead to almost equal growth in banks’ excess reserve holdings, and therefore to very little growth in either bank deposits or monetary aggregates. That is, IOER would have the same effect during the Fed’s rounds of QE as it had beforehand, when the Fed’s balance sheet was expanding, not as part of a deliberate monetary stimulus program, but as the incidental consequence of its emergency lending.

If it’s indeed true that “insanity is doing the same thing over and over again, but expecting different results,” then one might be excused for wondering whether, in expecting extra bank reserves to stimulate the economy after 2008, using the same
operating framework they relied upon to prevent extra reserves from stimulating the economy following Lehman’s collapse, Fed officials were playing with a full deck. In fact, despite an almost 4.5-fold increase in the monetary base between December 2008 and December 2014, bank deposits grew only about 60 percent, while the inflation rate broached the 2 percent mark only fleetingly. 29 As Figure IX.4 shows, the growth in M2 was just as disappointing 30

![FRED chart](image)

**Figure IX.4: The Monetary Base and M2, 2006-2015**

Although these outcomes took many commentators by surprise—including more than a few who feared that the Fed’s asset purchases would lead to high, if not hyper, inflation—they did so only because these commentators hadn’t grasped the implications of the Fed’s new operating framework. The simple truth was that, by switching to a floor

---

29 The Fed’s three rounds of Large Scale Asset Purchases have informally come to be known since as QE1, QE2, and QE3. QE1, which ran from December 2008 to June 2010, added $2.1 trillion, mainly in Mortgage-Backed Securities (MBS), to the Fed’s balance sheet. For QE2, which ran from November 2010 until June 2011, the Fed bought $600-billion worth of Treasury securities. QE3, finally, began in September 2012, and consisted of an open-ended program of securities purchases, starting with $40 billion in MBS per month, and supplemented, beginning in December 2012, with monthly purchases of another $45 billion in long-term Treasury securities. In all, between December 2008 and October 2014 the Fed purchased securities worth not quite $4 trillion, or about 4.5 times its total assets just prior to the crisis.

30 The Fed’s QE experience was foreshadowed, albeit on a much smaller scale, by the Reserve Bank of New Zealand’s mid-2006 decision to more than double that nation’s monetary base, raising it between then and December 2006 from NZ$6 billion to NZ$14 billion while concurrently establishing a floor system by increasing its deposit rate “by 5 basis points five times between July and October 2006 for a total increase of 25 basis points” (Koning 2016)—enough to guarantee that banks would sit on their new reserves. The (big) difference between this and the later U.S. episode was that the RBNZ wasn’t trying to achieve a macroeconomic stimulus. Instead it merely wished to enhance the liquidity of New Zealand’s banking system—a goal for which its policy was well-suited.
system in which excess reserves no longer carried a substantial opportunity cost, the Fed had dramatically reduced the potency of its asset purchases by crashing the money multiplier. As Figure IX.5 shows, the M2 base multiplier (the ratio of M2 to the monetary base), which had hovered between 8 and 9 prior to the implementation of IOER, has since fallen to half that value or less. Yet severe as that decline was, it was nothing compared to the astonishing decline in the bank reserves-to-deposits multiplier, which fell from an average value of about 750 in months prior to September 2008 to one of around ten from December 2008 onwards!\textsuperscript{31}

But Fed officials, far from being confused, were fully cognizant of these developments.\textsuperscript{32} They understood, furthermore, that they meant that Quantitative Easing, if it was to stimulate at all, could not do so by means of the usual multiplier-based monetary transmission mechanism. Indeed some, including Bernanke himself, went so

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure.png}
\caption{The Base-to-M2 Multiplier}
\end{figure}

\textsuperscript{31} To observe that the values of various money multipliers collapsed after September 2008 is not at all to suggest, as Martin Wolf (2014) and some other pundits have, that the very notion of a money multiplier is a “myth.” Of course there are no absolutely constant money multipliers. But while textbooks may occasionally assume fixed multipliers for convenience’s sake, competent monetary economists have long understood that multipliers can vary, and sometimes dramatically. Nor (despite what Mr. Wolf says) were matters any different in the days of the gold standard: in the U.S. during that time, for example, the variability of the base multiplier was a notorious source of concern.

It is also not correct to suggest, as some other writings (e.g., Martin McAndrews and Skeie 2013, p. 1) have, that the existence of a substantial multiplier depends upon the presence of minimum legal reserve requirements. On the contrary: so long as banks have a well-defined demand for either clearing balances or till money, eliminating reserve requirements, far from rendering the multiplier indeterminate, serves only to raise its value, other things equal. So long as required reserve and equilibrium currency-deposit ratios are less than one, the existence of a positive opportunity cost of reserve holding is both a necessary and a sufficient condition for the existence of a positive money multiplier.

\textsuperscript{32} See, for example, Williams (2012).
far as to object to the expression “Quantitative Easing” because it suggested, misleadingly, that the Fed regarded LSAPs as a means for expanding the quantity of money, and for giving a boost thereby to spending, prices, and employment. “The theory behind quantitative easing,” Bernanke (Board of Governors 2008c, p. 24) explained at the FOMC’s December 2008 meeting, is

that providing enormous amounts of very cheap liquidity to banks…would encourage them to lend and that lending, in turn, would increase the broader measures of the money supply, which in turn would raise prices and stimulate asset prices, and so on, and that would suffice to stimulate the economy. ... I think that the verdict on quantitative easing is fairly negative. It didn’t seem to have a great deal of effect, mostly because banks would not lend out the reserves that they were holding.

Just why “banks would not lend out the reserves that they were holding” Bernanke didn’t say; nor did he so much as hint at the possibility that it had something to do with the Fed’s decision to pay interest on those reserves—a policy originally adopted, let’s not forget, to keep banks from “increasing the overall supply of credit to the economy” despite Fed-engineered additions to the total stock of bank reserves (Walter and Courtois 2009).

In any event he and other Fed officials hoped that the Fed’s asset purchases might instead influence the real economy through other channels. In particular, they appealed to the existence of a “portfolio balance” channel, in which changes in nominal quantities, and in bank lending especially, played no essential part. Instead, the Fed’s asset purchases were supposed to boost real economic activity by altering relative asset prices. In particular, swapping bank reserves for long-term securities was expected to promote investment by lowering long-term interest rates.

But whether there really is such a thing as a portfolio balance channel is a matter of considerable controversy. Just before he left the Fed Bernanke, when asked how
confident he was in QE’s effectiveness, famously replied that “The problem with QE is it works in practice, but it doesn’t work in theory” (quoted in Harding 2014). Though said in jest, there was more than a little truth in Bernanke’s remark—or in the last part of it at any rate. And Bernanke knew it. As a 2014 Financial Times article explains, according to theory that prevailed in the years before the crisis,33 so long as banks themselves are indifferent between holding new excess reserves and trading them for other assets, as they would be at the zero lower bound in the absence of IOER, and as they are if reserves bear interest at or above the going market rate, the Fed’s own asset purchases should have no effects. All that happens is the central bank swaps one kind of government debt—money—for another kind of government debt, in the form of a long-term Treasury bond. That can only make any difference if investors have a strong preference for one kind of debt over the other (Harding 2014).

For the portfolio-balance channel to be relevant, it had to be the case, as Bernanke himself explained in his 2012 Jackson Hole speech, that “different classes of financial assets are not perfect substitutes in investors’ portfolios” (Bernanke 2012; my emphasis).

Stephen Williamson (2017b, pp. 11-12), then of the St. Louis Fed, supplied what I believe to be the best assessment of dubious theoretical presuppositions of Quantitative Easing:

A central bank is a financial intermediary. It borrows from a large set of people—those who hold the central bank’s primary liabilities, i.e., currency and reserves. And the central bank lends to the government, private financial institutions and sometimes to private consumers...

Like private financial intermediaries, central banks transform assets in terms of maturity, liquidity, risk and rate of return. Therefore, the ability of a central bank

---

33In particular, Eggertson and Woodward (2003).
to affect economic outcomes in a good way depends on its having an advantage relative to the private sector in intermediating assets. ... 

When QE is conducted in a system flush with reserves, the central bank is typically transforming long-maturity assets into short-maturity reserves. The key question, if we compare this to how conventional monetary policy works, is what advantage the central bank might be exploiting in conducting such a transformation. That is not clear. ... Therefore, from financial intermediation theory, it is not clear that QE should have any effect and it might actually be detrimental to the efficiency of the financial system.

But did QE at least work in practice, as Bernanke claimed it did? In the aforementioned Jackson Hole speech, Bernanke went on to refer to statistical evidence that the Fed’s strategy had succeeded. But many other economists find this same evidence far from convincing. Williamson, for one, considers it “pretty sketchy”:

For the most part, the empirical work consists of event studies—isolate an announcement window for a policy change, then look for movements in asset prices in response. There's also some regression evidence, but essentially nothing (as far as I know) in terms of structural econometric work, i.e. work that is explicit about the theory in a way that allows us to quantify the effects (Williamson 2017a; see also Williamson 2012 and Williamson 2017b pp. 12-14).

The positive findings, furthermore, generally concern QE’s effects on bond yields only, rather than on more important macroeconomic variables, such as inflation and unemployment. As Mirco Balatti and his coauthors (2016, p. 3) quite properly observe, to conclude that QE was “effective” merely because it altered bond yields is to toy with the usual meaning of monetary policy effectiveness, by conflating a policy’s success in influencing an intermediate policy target with its success in achieving ultimate policy goals. There are, after all, good reasons, as Charles Goodhart (2017) argues, for suspecting
that the real implications of any interest rate cut will depend on just how that cut has been achieved:

One does not have to be a full-blown Monetarist... (and I am not), to feel that a condition whereby a 100 b.p. cut in interest rates is accompanied by a 5% increase in broad money is likely to be much more expansionary than when the same cut in interest rates is matched by no such increase.

Since “[many] persons primarily access the financial scene via intermediaries,” Goodhart continues, then “[i]f the transmission mechanism, via banks and other non-bank financial intermediaries, is seriously impaired, then so will be the efficacy of a given initial interest rate policy change, a point that most formal models miss” (ibid.).

According to Balatti et al. (2016, p. 5), while QE did indeed lower interest rates, it otherwise “struggled to propel the macroeconomy.” St. Louis Fed economist Yi Wen (2014, pp. 1-2) likewise observes, with particular regard to the Fed’s large-scale MBS purchases (“credit easing”), that while it “has had significant effects on lowering the real interest rate of private credit and raising asset prices...the picture on the real side of the economy looks gloomy, making it difficult to claim victory for [credit easing] despite 5 years in the making.” According to Wen’s own analysis (ibid., p. 4), even if they’d been permanent the Fed’s MBS purchases “would have had little effect on aggregate output and employment.” Even some top Fed officials who oversaw the Fed’s QE operations were far from sanguine about their success. “I think it’s fair to say,” former Vice Chair Donald Kohn remarked, that “although [LSAPs] were effective to some extent, people—even the Fed—were somewhat disappointed. It’s been a slow recovery from a very deep recession” (quoted in Harding 2014).34

A very recent study has, finally, cast doubt even on the limited claim the Quantitative Easing reduced bond yields substantially. According to Hamilton, Harris, and West (2018), although earlier event studies had informed a consensus view that QE

---

34 For a thorough evaluation of QE, see Thornton (2015).
lowered yields on 10-year Treasury bonds by about 100 basis points (Gagnon 2016), controlling for an additional event suggests instead that, “in marked contrast to the effects of conventional policy in the pre-2008 era,”—that is, before the Fed switched to a floor system—the effects of the Fed’s Large-Scale Asset Purchases were neither large nor persistent (ibid, p. 39).

**IX. c. Stimulus without IOER?**

The most important question concerning the Fed’s approach to post-crisis stimulus is, not whether it was at all successful, but whether another approach might have been better. In particular, what would have happened had the Fed dispensed with IOER, and the floor system it supported, while still expanding its balance sheet?

It happens that Fed officials themselves considered this very question as the Fed was deciding, earlier in the summer of 2010, whether to renew the QE1 asset purchases it had tentatively ended that June. As Ben Bernanke (2010b) reported in his Jackson Hole speech that August, having contemplated “reducing the IOER rate to, say, ten basis points or even to zero” as one of several alternatives to having the Fed buy more assets, he and his colleagues concluded that

On the margin, a reduction in the IOER rate would provide banks with an incentive to increase their lending to nonfinancial borrowers or to participants in short-term money markets, reducing short-term interest rates further and possibly leading to some expansion in money and credit aggregates. However, under current circumstances, the effect of reducing the IOER rate on financial conditions in isolation would likely be relatively small. The federal funds rate is currently averaging between 15 and 20 basis points and would almost certainly remain positive after the reduction in the IOER rate. Cutting the IOER rate even to zero would be unlikely therefore to reduce the federal funds rate by more than 10 to 15 basis points. The effect on longer-term rates would probably be even less, although that effect would depend in part on the signal that market participants took from the action about the likely future course of policy.
These conclusions are striking for several reasons. They imply, first of all, that Bernanke and his colleagues no longer believed, if they ever did, that IOER alone stood in the way of having the fed funds rate decline to zero. Second, given the prevailing circumstances, it is difficult to see why the Fed would resist a potentially helpful IOER rate reduction just because the anticipated benefits would be “relatively small.” Third, recalling VanHoose and Dutkowsky’s (2017, p. 11) calculation that “banks would not have accumulated reserves so long as the effective federal funds rate itself remained at or above 6 basis points,” the anticipated decline in the fed funds rate would almost certainly have left it high enough, with IOER = 0, to sponsor a switch back to the pre-2008 zero excess reserves regime, the stimulus effect of which would have been anything but “relatively small.” Instead of considering the “expansion in money and credit aggregates” to which the elimination of IOER might lead, Bernanke and his colleagues failed to look beyond that step’s more immediate interest-rate effects, implicitly treating a low- or no-IOER alternative as just another way of achieving a portfolio balance effect, rather than as a way to put the Fed’s traditional monetary transmission mechanism back in gear. By thinking so they could hardly avoid dramatically underestimating the alternative policy’s potential benefits.\textsuperscript{35}

Despite Bernanke’s doubts, the possibility of lowering the IOER rate to zero remained a live one for some time. About a year after Bernanke’s Jackson Hole remarks, in response to more bleak news about the pace of the recovery, the FOMC again considered lowering the IOER rate as one of three alternatives, the others being another round of QE and what would become known as “Operation Twist,” for easing policy (Board of Governors 2011). Eventually the committee voted to stand pat until September 2011, when it decided to go ahead with Operation Twist. During the FOMC’s July 31-August 1, 2012 meeting, held after the ECB had cut its own deposit rate to zero, some members again raised the possibility of lowering the IOER rate, but the possibility was

\textsuperscript{35} At least one Federal Reserve Board economist, Joseph Gagnon (2010), thought that the IOER rate should be lowered nonetheless. In a July 2010 blog post he wrote that “the Fed should lower the interest rate it pays on bank reserves to zero. This is a small step, as the current rate is only 0.25 percent, but there is no reason to pay banks more than the rate paid by the closest substitute, short-term Treasury bills.”
again rejected (Board of Governors 2012b, p. 8). Although concerns such as those Bernanke had expressed earlier may have influenced these decisions, a more immediate concern—and one Wall Street itself voiced with great vigor—was that cutting the IOER rate would harm money market funds by lowering yields on Treasury bills and other short-dated securities of the sort upon which those funds relied (ibid; Robb 2012). To the extent that the concern was a valid one, it was also short-sighted, for although in attempting to dispose of excess reserves by acquiring other short-term assets banks might initially cause the yield on short-dated securities to decline, the revival of bank lending would ultimately have contributed to a general increase in spending and inflation, thereby raising equilibrium nominal interest rates.

Whatever the consequences of a lowered IOER rate for the money market funds, the relevant question for policymakers should have been whether lowering the IOER rate would have helped the U.S. economy as a whole. Finadium’s Jonathan Cooper (2012) believes it would have, and that for that reason the Fed made a mistake by not following the example the ECB set by cutting its own deposit rate in June 2012:

If IOER went back to what it [sic] was originally intended—a rate pegged at below where the Fed Funds rate actually cleared in the market—it might be a first step toward banks thinking more clearly about how to put their cash to work, restore Fed Funds credibility as a market benchmark, and allow the Fed to start to get out of the box they’ve painted themselves into.

To see why, consider again our diagrammatic representation of the influence of IOER on bank lending, reproduced here as Figure IX.5:
The Fed’s assessment mainly takes account of the movement along a given, downward-sloping loan demand schedule associated with a decline in the IOER rate. But recall that that movement represents an equilibrium only assuming that the Fed withdraws reserves from the banking system to an extent equal to the decline in the quantity of reserves demanded following the rate reduction. Otherwise the banks will dispose of those same unwanted reserves by exchanging them for other, interest-yielding assets; and they will continue doing so until deposits have grown to the point at which the quantity of reserves demanded is again equal to the quantity supplied. The reserve-deposit multiplier would, in other words, be reinvigorated, at least to some extent; and that revival would in turn mean that the overall volume of bank lending, instead of merely increasing to the extent implied by a movement down a fixed loan demand schedule, would increase further by virtue of a general increase in nominal magnitudes, entailing a corresponding rightward-shift in the loan demand schedule.

To get some idea of how much even a partial revival of the money multiplier would have mattered, consider that, over the decades prior to Lehman’s failure, every dollar of base money supported between 5 and 8 times as many dollars of bank lending, the higher figure having been reached just before the crash. In contrast, as Figure IX.6 shows, by June 2010, mainly thanks to IOER, the ratio had fallen to 3.38. Consequently, even assuming, very conservatively, a post-Lehman base-money-to-loans multiplier of 5, the

Figure IX.5: Bank Lending and IOER
elimination of interest payments on reserves in the summer of 2010 would, other things equal, have raised the equilibrium value of bank lending by more than $3.2 trillion.

Figure IX.6: The Monetary Base and Bank Deposits and Loans, 2003-2011

Besides underestimating the extent to which ending IOER might boost bank lending, Bernanke (2010b) and his colleagues worried that it could lead short-term money markets such as the federal funds market to become much less liquid, as near-zero returns might induce many participants and market-makers to exit. In normal times the Fed relies heavily on a well-functioning federal funds market to implement monetary policy, so we would want to be careful not to do permanent damage to that market.

Here the Fed’s reasoning was not only incorrect, but disingenuous. If anything prevented the fed funds market from functioning as it had “in normal times,” above-market IOER was it! Although it’s true that reducing the IOER rate to zero would have eliminated the arbitrage opportunity that was responsible for most of the fed funds lending that occurred while the Fed’s floor system was in place, as banks disposed of excess reserves they no longer wished to hold, bringing their holdings back to minimal levels, “normal” interbank lending would eventually resume. In other words, what Fed officials were weighing as potential “damage” to the fed funds market was but a stage in the restoration of that market’s robust health.
IX. d. Stimulus at the Zero Bound?

And suppose that the “neutral” or “natural” value of the federal funds rate had in fact been negative, so that, absent IOER, the effective fed funds rate would have fallen to zero, Fed officials contrary estimates notwithstanding. Would that in turn have meant the end of any potential stimulus, save what might be achieved by means of large-scale asset purchases?

Had the Fed’s choice been between resorting to such purchases and following a conventional Taylor Rule, as it appeared to do during much of the Great Moderation, the answer might be yes. However, even allowing that some sort of interest-rate feedback rule was the only practical alternative, a number of studies employing several different models suggest that relatively modest departures from the conventional Taylor Rule would have allowed the Fed to achieve its macroeconomic objectives, despite the zero lower bound, and to have done so without dramatically increasing the size of its balance sheet. In every case the solutions consist of very similar “history-dependent” modifications to the Taylor Rule, with such forward guidance as those modifications imply.

In one of the earlier studies, David Reifschneider and John Williams (2000, p. 957) showed, using the Fed’s own FRB/US model—a large-scale open-economy model that’s been at use there since 1996—that the Fed might overcome the zero bound problem using a rule that makes up for periods in which the Taylor Rule calls for a negative policy rate by systematically keeping the funds rate at zero for an extended period to compensate for a “backlog” of unachieved, negative rates.36 As an example, Reifschneider and Williams offer a situation in which, according to the standard Taylor Rule, the Fed would have to reduce its policy rate to minus 2 percent in order to avoid falling short of

---

36 That Reifschneider and Williams reach this conclusion using the FRB/US model is of special relevance, since that model, besides being the Fed’s own standard general equilibrium model of the U.S. economy, allows only for a very orthodox monetary transmission mechanism, in which ordinary open market operations are the Fed’s sole instrument of monetary control, and those operations influence broader money aggregates and inflation through “the joint interaction of the money demand equation and the reserves multiplier” (Reifschneider and Williams 2000, p. 947 n. 9)
its 2 percent inflation target. In that case, an alternative solution (shown in Figure IX.7) might be for the Fed to commit to delay raising its policy target above zero until some time after the Taylor Rule itself would can for such an increase.

Figure IX. 7: Standard and Modified Taylor Rule Rate Settings for Binding Zero Bound (Reproduced from Reifschneider and Williams 2000, p. 958)

In their theoretical study of the optimal policy-rate path when the zero lower bound is binding Taehun Jung and his coauthors (2005) reach conclusions quite similar to those of Reifschneider and Williams. That optimal path, they conclude, “is characterized by monetary policy inertia, in the sense that a zero interest rate policy is continued for a while even after the natural rate of interest becomes positive” (ibid, pp. 824-5). A central bank that sticks to such an “augmented” Taylor Rule

is able to achieve higher expected inflation and lower long-term nominal interest rates in the periods when the natural rate of interest significantly deviates from a steady-state level, thereby stimulating aggregate demand in these [zero bound] periods. That is as if the central bank “borrows” future monetary easing in the periods when current monetary policy easing is already exhausted (ibid, p. 825).

Regarding a few other studies in the same spirit: using a New Keynesian model, Fernando Duarte and Anna Zabai (2015) show that the zero lower bound problem can be overcome by having the period of time for which the central bank commits to keep its
policy rate “pegged at zero” depend positively on the extent to which it falls short of its inflation target. Nor need the extra zero-bound time commitment be very long to achieve its purpose. For example, if inflation falls short of its target by two percentage points, their model calls for a commitment to stay at zero for 2.8 years, as opposed to the 2.6 years that would be called for by a standard Taylor Rule. Siddhartha Chattopadhyay and Betty Daniel (2017) likewise show that a central bank can always achieve a desired degree of stimulus at the zero lower bound by adhering to a Taylor Rule modified to allow for both a time-varying inflation target and an optimally-chosen zero-bound exit date.

Finally, Hu McCulloch (2018, pp. 12-15) supplies an informal version of the same basic argument. Suppose, he says, that inflation is running at zero, or 2 percentage points below the Fed’s inflation target, and that, according to a “benchmark” Taylor Rule, the Fed’s optimal nominal funds rate target is minus 1 percent, implying an initial real fed funds rate that’s 3 percentage points below the assumed natural real rate ($r^*$) of 2 percent. Because it’s constrained by the zero lower bound, the Fed, were it to simply commit to a zero rate for the usual 6-month rate-setting period, would achieve only 2/3 of the needed stimulus. However, because “[l]owering the nominal interest rate γ(mo) on loans of maturity $mo$ by $Δi$, while holding forward rates beyond $mo$ constant, reduces the cost of borrowing to any maturity beyond $mo$ by $moΔi,” the Fed might still achieve the necessary stimulus by promising to keep the fed funds rate pegged at zero for at least 9 weeks. “Doing this,” McCulloch adds,

with no direct disturbance to forward rates beyond 9 weeks would require the Open Market Desk to peg the interest rate on T-bills maturing within 9 weeks of the current FOMC meeting to 0, and to hold them there until the next FOMC meeting. At that time, the FOMC would then be free to continue the stimulus by moving the peg out another 6 weeks, or to alter the strength of the stimulus in either direction (ibid.).

Though studies like these can hardly be said to leave us in no doubt concerning whether the Fed might have allowed reserve growth to drive the fed funds rate all the way
to its zero lower bound, while nevertheless retaining its ability to stimulate the economy, they at least suggest that the possibility shouldn’t have been dismissed lightly. They also suggest that the Fed’s chosen alternative—a policy rate held above zero by means of interest on reserves—was more likely counterproductive than not, for that policy meant having a gap between the Fed’s chosen policy target and its optimal (negative) policy rate setting greater than the one due to the zero lower bound itself, and a correspondingly greater dependence upon offsetting forward guidance.

IX. e. Positive, Zero, or Negative?

Although we’ve been assuming that, when short-term natural interest rates sink into negative territory, policy rates cannot follow them there, experience shows that this assumption isn’t always valid. Indeed, not long after Bernanke and his fellow FOMC members had decided against lowering the Fed’s IOER rate to zero, central bankers elsewhere had begun to contemplate, not merely ceasing to pay above-zero interest on banks’ reserve balances, but switching to negative interest rates. On paper, at least, this was the “obvious” solution to the problem of a zero lower bound, though it was so only in the sense that assuming one has a can opener is the obvious solution to the problem of opening a tin can. In practice the technical and legal difficulties involved in making such a switch were far from trivial. Yet many central banks, starting with Denmark’s Nationalbank, have managed to do it.\footnote{Denmark’s example was quickly followed by the ECB. Eventually the central banks of Switzerland, Sweden, Norway, Japan, Hungary, and Bulgaria also implemented negative deposit rates.} Among them, the Danish and Swiss central banks plunged deepest, each having for a time set its deposit rate at minus 75 basis points, or 100 basis points below the Fed’s lowest IOER rate.

Suppose that, instead of sticking to its positive-IOER floor system, the Fed had also chosen to combat the Great Recession by implementing a negative interest rate policy? Laurence Ball, Joseph Gagnon, Patrick Honohan and Signe Krogstrup (2016) performed a counterfactual simulation for the period 2008 through 2015 in which they assumed that
the Fed’s policy rate, instead of following its actual path, was consistently set according to the Taylor Rule, even when that rule called for a negative rate.

According to their results, by early 2009 the Fed would have lowered the fed funds rate to about minus 6 percent, or 625 basis points lower than the actual IOER rate. That lower rate would in turn have caused unemployment to decline more rapidly, reducing the unemployment gap to just one percentage point, or half its actual value, by 2012. Inflation, on the other hand, would have been about a percentage point higher, briefly exceeding, instead of falling short of, its 2-percent target. “Because a faster recovery means that monetary conditions can normalize more quickly,” the authors observe, their counterfactual federal funds rate “becomes positive in 2011, more than four years before the Fed actually increased the rate” (ibid, p. 11). Ball et al. (ibid, p. 23) also ask what difference it would have made had the Fed been limited to a negative rate of 2 percent or less. They find in that case that the fed funds rate, having fallen to the new minus 2 percent lower bound, would have remained there for just 10 quarters, and that “the unemployment rate would have been about 0.7 percentage points lower on average starting in 2010, for a cumulative reduction in the unemployment gap of about 4 percentage points.” Inflation would also have been higher, though only slightly.

Such counterfactuals must be taken with a grain of salt, of course. Ball et al. never suggest that the Fed could actually have implemented a policy rate of minus 2 percent, let alone minus 6. Furthermore it is far from clear that, by lowering its policy rate from zero to, say, minus 50 basis points, a central bank might achieve a degree of stimulus comparable to what it would achieve by lowering its policy rate from 50 basis points to zero. That commercial banks may find it difficult to charge negative deposit rates, so as to preserve their net interest margins as the IOER and wholesale rates fall below zero, and that their attempts to do so might only inspire the public to swap bank deposits for cash,
are only two of the more well-discussed difficulties that can limit NIRP’s overall effectiveness.\footnote{Eggertsson et al. (2017) claim that, because it can chip away at bank profits, NIRP may in some cases prove to be contractionary. Studies of Europe’s experience with NIRP (Jobst and Lin 2016; Dell’Ariccia et al. 2017) conclude that, while it may have had a slight adverse effect on bank profits, it has also contributed to modest credit and spending growth.}

Still the point remains that, if a negative IOER rate was “optimal” in theory—that is, if it would have done the trick of hastening the recovery provided banks could pass the higher reserve tax on to their own depositors—a positive IOER rate can only have slowed the recovery down. That the Fed chose to press ahead with its positive-IOER floor system even as other central banks found it worthwhile to experiment with negative policy rates ought at very least to raise doubts about Fed’s choice: if one of these opposite central bank gambits was prudent, the other almost certainly wasn’t.

\textit{IX. f. Canada’s Counterexample}

At the time of the crisis the Bank of Canada operated a symmetrical corridor system, in which the Bank of Canada’s lending rate (“Bank rate”) served as the channel’s upper bound, while the “deposit rate” it paid on banks’ overnight balances—the counterpart of the Fed’s IOER rate—served as its lower bound. Until April 21, 2009, these upper and lower bounds were respectively set at 25 basis points above, and 25 points below, the Bank’s chosen overnight lending rate target. The Bank’s IOER rate was therefore a below-market rate, by design as well as in fact. As Canadian banks weren’t subject to any minimum reserve requirements, this arrangement encouraged them to keep their overnight reserves at a bare minimum, typically equal to about C $25 million.

Between April 21, 2009 and June 1, 2010, in response to Canada’s worsening recession, the Bank of Canada switched briefly to a floor system, by setting both its target and its deposit rate at 25 basis points, thereby making banks indifferent between holding overnight balances and lending them. To drive the overnight rate to the floor, it provided banks with an additional C $3 billion in excess reserves, which they duly kept at the
Bank’s standing deposit facility. The Bank also promised to keep the rate at 25 basis points until the second quarter of 2010, with the aim of achieving its inflation target, thereby making effective use of “forward guidance” to influence the public’s interest-rate expectations (Woodford 2012, pp. 12-16). The policy having worked as intended, on June 1st the Bank re-established its normal operating corridor by raising its overnight rate target to 50 basis points, while keeping its deposit rate 25 basis points (Bank of Canada 2010).

Had Fed officials been right in thinking that continued reserve creation, coupled with above-market IOER, was a more reliable means for stimulating economic activity than ending banks’ extraordinary demand for reserves would have been, the Bank of Canada’s chosen response to the crisis, with its meager and temporary boost to banks’ excess reserve holdings, must surely have been less effective than the massive and sustained increase in U.S. banks’ excess reserve holdings overseen by the Fed. Yet, as the figures below, reproduced from Stephen Williamson’s (2017a) blog, show, if anything the opposite was true: as Figures IX.9 and IX.10 show, Canada’s price level and real GDP each recovered somewhat more rapidly than their U.S. counterparts from their spring 2009 nadirs. “As an econometrician once told me,” Williamson wryly observes in commenting on them,

if I can’t see it, it’s probably not there. Sure, since Canada is small and is highly integrated with the US economically, Fed policy will matter for Canadian
economic performance. But, if QE were so important, the fact that the US did it and Canada did not should make some observable difference for relative performance (ibid.).

Figures IX.9 and IX.10: Canadian and US CPIs (Left) and Real GDP, 2008-2017

IX. g. An Overtightening Bias?

Finally, the Fed’s persistent post-crisis failure to reach its desired inflation target raises further doubts concerning the adequacy of its new operating system, and especially so since January 2012, when, for the sake of keeping the public’s inflation expectations “firmly anchored,” the Fed announced an explicit inflation target, consisting of a 2 percent annual increase in the Personal Consumption Expenditure (PCE) index (Board of Governors 2012a). In making that announcement, the Board of Governors declared that “the inflation rate over the longer run is primarily determined by monetary policy” (ibid.). That was certainly true under the Fed’s traditional operating system, as is evident in the pre-crisis behavior of the PCE index, as shown in Figure IX.11 below. During that time, for better or worse, the Fed had no difficulty maintaining a PCE inflation rate just a little in excess of 2%, which was then, according to many, the Fed’s implicit inflation target.39 In contrast, since it announced its explicit PCE target, with its new stuck-in-neutral operating system in place, the Fed has failed to reach that target in every quarter save that

39 “For better or worse” because maintaining a steady inflation target at a time of rapid productivity growth meant tolerating unusually rapid NGDP growth, which may have contributed to the subprime bubble. See Borio and Lowe (2002).
of the announcement itself—and has done so despite adding over one trillion dollars to banks’ reserve balances!

As the *New York Times* reported recently, although “the direct cost of mildly undershooting the Fed’s inflation target is low,”

What is worrisome is not the direct damage, but the fact that the Fed has missed its (arbitrary) 2 percent target in the same direction—undershooting—year after year... . That in turn implies that the low-growth, low-inflation, low-interest rate economy since 2008 isn’t going anywhere. This would prove especially damaging if the economy ran into some negative shock; a lack of Fed credibility could leave it less able to prevent a recession (Irwin 2017).

What has the floor system to do with the Fed’s difficulty in reaching its inflation target? Of course, in any sort of operating system an overly-high rate policy rate setting will mean overtightening, with nominal variables falling short of their targets. The problem with a floor system is that it places an extra constraint on policy rate settings, consisting of the requirement that the IOER rate be at least as high as the net return on other short-term assets. We’ve seen, furthermore, that even very modest changes in the relation between the Fed’s IOER rate and other interest rates can suffice to trigger a switch, whether intentional or not, from a floor regime to a corridor system, with the
dramatic unleashing of the reserve multiplier such a switch entails. For that reason, Donald Dutkowsky and David VanHoose explain (2017, p. 11), it would only be prudent for the Fed to endeavor to keep its IOER rate well above any level that might risk an unplanned regime change.

Because the floor-system-preserving IOER constraint takes the form of an inequality, it means that a central bank that’s determined to keep a floor system in place is especially likely to err on the side of monetary overtightening. Suppose, for example, that, at some initial, general level of market rates, the Fed’s floor-system IOER setting is allowing it to just hit its inflation target, and that by that criterion at least rates, including the IOER rate, can be said to be at their “neutral” or “natural” levels. Now suppose that the neutral real rates increase by 10 percent. To continue hitting its inflation target, the Fed must also raise the IOER rate by 10 percent, while to maintain its floor system it must raise the IOER rate by at least ten percent. In contrast, if neutral rates decline 10 percent, the Fed can maintain its floor system only if it lowers its IOER by no more than 10 percent: anything lower would inspire banks to unload their excess reserves, reviving the money multiplier. It’s owing to this asymmetry that a floor system may be said to tolerate, if not to invite, excessively tight, but not excessively loose, policy.

Has the Fed’s determination to keep its floor system intact caused it to err in recent years on the side of excessively tight money, by raising its policy rates more aggressively than underlying neutral rate increases warranted? The possibility is certainly worth considering. Increases in the IOER rate starting with that of December 2015 appear, at least superficially, to have been more consistent with the objective of keeping the Fed’s floor system in place than with that of achieving its macroeconomic targets. In any event the risk that a commitment to it might lead to overly tight policy in the future should be considered another shortcoming of the Fed’s post-crisis operating system.
X. The Floor System and Credit Allocation

X. a. Central banking versus Commercial Banking

As we’ve seen, although the total quantity of Fed reserve balances is mainly a function of the size of the Fed’s balance sheet, the quantity of excess reserves banks hold ultimately depends on banks’ demand for such reserves, as influenced by their yield relative to other assets. Until it switched to a floor system, the Fed discouraged banks from holding excess reserves, by confronting them with a substantial opportunity cost of doing so. The new system, in contrast, encourages banks to hold unlimited quantities of excess reserves and to lend a correspondingly large part of the savings they secure from depositors to it. As Hu McCulloch (2018, p. 3) puts it,

The Fed in effect is now acting as a huge financial intermediary, borrowing reserve deposits from the banks at interest and lending them back to homeowners as mortgages or by transforming the maturity of the National Debt.

The Fed’s increased role in financial intermediation is evident in Figure X.1, which shows that bank reserves, which until the recent crisis made up only a fraction of a percent of total bank deposits, are now equal to a fifth of those deposits. Bank lending to businesses, farmers and consumers, on the other hand, has gone from roughly matching total bank deposits to being equal to only four-fifths of those deposits. The Fed has thus made itself responsible, not merely for regulating the nominal scale of deposit-based financial intermediation in the U.S. economy, but for actually disposing of a substantial share of the public’s savings. As the chart below shows, relative to the assets held by the entire U.S. commercial banking system, the Fed’s holdings are now four times what they were before the crisis.
Were the Fed itself just another commercial bank, it might be expected to employ the public’s savings at least as efficiently as commercial banks themselves might, by directing them to uses offering relatively high risk-adjusted returns. But the Fed is not a commercial but a central bank. As such it was never intended to act as an efficient financial intermediary, whether by directly competing with commercial banks, or by having them serve as mere agents to it, as they do when they hold substantial excess reserves. Unlike commercial banks a central bank’s purpose is to secure macroeconomic stability and to otherwise attend to the interests of the public at large, rather than to those of its nominal owners.

The Fed’s unique responsibilities have as their counterpart unique operating principles that differ greatly from those appropriate to commercial banks, including guidelines concerning both the sort of assets it should invest in, and the extent of its overall involvement in credit allocation. A relatively recent statement of these guidelines can be found in a 2002 Federal Reserve System Study Group report on “Alternative Instruments for System Operations” (Board of Governors 2002). Among other things the report states that as a public entity the Fed should “manage its portfolio to be adequately

---

40 Ideally, this would mean high risk-adjusted returns. That implicit and explicit guarantees encourage at least some actual commercial banks to engage in excessively risky lending is notorious.
compensated for risks” while also maintaining “sufficient liquidity in its portfolio to conduct potentially large actions on short notice.”

Until 1966 these principles were met by limiting the Fed’s open-market purchases to “Treasuries only,” meaning short-term Treasury securities. However, in that year Congress amended the Federal Reserve Act to temporarily allow the Fed to purchase any fully-guaranteed agency securities, and in 1968 that change was made permanent (Haltom and Sharp 2014).

The Fed was also supposed to “structure its portfolio and undertake its activities so as to minimize their effect on relative asset values and credit allocation within the private sector (Board of Governors 2002).” This last rule, the same report continues, is consistent with well-supported doctrines in the economics literature: In general, market price mechanisms allocate resources most effectively when undistorted by government actions, and market-directed resource allocation fosters long-run economic growth. The truth of these doctrines also has been borne out by much hard experience, both domestic and international, with varying levels of governmental intervention in the market process (p. 1-2).

According to the FOMC’s original guidelines concerning them, even the Fed’s agency security purchases were “not designed to support individual sectors of the market or to channel funds into issues of particular agencies” (Board of Governors 1971, p. 997).42

X. b. IOER and Financial Repression

Because the Fed’s own portfolio choices are limited, and especially because it doesn’t extend credit to nonfinancial firms or individuals, it can’t be expected to employ

---

42 According to Renee Haltom and Robert Sharp (2014, pp. 6–7), during the December 2008 FOMC meeting held after QE1 had been announced, then Richmond Fed President Jeffrey Lacker observed that that plan appeared inconsistent with the guidelines in question, and particularly with the Fed’s press release stating that the purchases were intended “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.” In January 2009, when the FOMC voted to suspend the guidelines indefinitely, Lacker alone dissented (Appelbaum 2013).
savings as efficiently or productively as commercial banks can. For that reason it is only reasonable that it should be expected to intrude as little as possible on “market-directed resource allocation” and, specifically, that it should avoid having banks hold unnecessarily large balances with it. Indeed, those central banks that do otherwise are generally condemned for engaging in what economists call “financial repression,” meaning practices that “prevent an economy’s financial intermediaries...from functioning at their full capacity,” thereby interfering with the efficient allocation of credit and impairing economic growth (Ito 2009, p. 430).43

Yet, as we’ve seen, by paying IOER at above-market rates while generating trillions of dollars in additional reserve balances, the Fed has increased the share of Fed-directed resource allocation, while reducing that of market-directed resource allocation by a corresponding amount. Instead of being directed by private-sector lenders to (mainly) private-sector borrowers, the resources represented by commercial banks’ excess reserve balances have instead been directed by the Fed towards those entities whose securities it purchased during several rounds of Quantitative Easing.

Some may wonder whether paying banks to accumulate excess reserves can really have the same, oppressive effects as imposing high reserve requirements might. If holding reserves pays more than other uses of funds, then isn’t it also efficient for banks to hold reserves instead of acquiring other assets? The answer is that it would certainly be efficient if the Fed’s relatively high IOER rates reflected its own capacity to employ funds more productively than private-market lenders. Instead the Fed’s ability to pay above-market IOER rates has been due, not to its being an unusually efficient intermediary, but to the seigniorage it earns on its non-interest-bearing notes and on non-interest-bearing

43 According to Ito (2009, p. 430), common examples of financially repressive policies include “interest rate ceilings, liquidity ratio requirements, high bank reserve requirements, capital controls, restrictions on market entry into the financial sector, credit ceilings or restrictions on directions of credit allocation, and government ownership or domination of banks.”

While high minimum reserve requirements and the use of relatively high IOER rates to induce banks to accumulate excess reserves both alter the direction of credit allocation, high reserve requirements also tend to enhance governments’ seigniorage revenues, while high IOER rates may not serve that purpose.
balances kept with it, which it can use to cross-subsidize bank reserves. Furthermore, because the Fed doesn’t practice mark-to-market accounting, it doesn’t have to provide for unrealized portfolio losses. Consequently, it is able to finance relatively high IOER rates in part by assuming greater risks, including the substantial duration risk it took on by acquiring long-term Treasury and mortgage-backed securities. By using those high IOER rates to compete with private-sector borrowers for commercial bank funds, the Fed alters both the scope and the direction of bank-assisted financial intermediation: while it raises somewhat the overall demand for bank deposits, and hence the overall extent of bank-assisted intermediation, by reducing the implicit tax on bank reserves, and thereby enhancing banks’ interest revenue, it also grabs a substantially larger share of the intermediated funds for itself.

To gain a more complete appreciation of the real consequences of above-market IOER rates on the one hand and high mandatory reserve requirements on the other, suppose, first, that instead of paying banks to hoard reserves the Fed achieved a similarly high reserve ratio by imposing a continuously-enforced 20-percent reserve requirement against all commercial bank deposits. Taken alone that step would lead to a severe contraction in nominal bank lending and bank deposits and, ultimately, to a corresponding decline in the price level. In the resulting equilibrium, the Fed’s real asset holdings would have grown, in both absolute terms and relative to commercial bank assets, to roughly the same extent as has happened in fact, though in a manner that would leave no doubt concerning the “repression” involved, consisting of a reallocation of savings from commercial banks to the Fed, and from commercial bank lending to bolstering the markets for Fed-favored securities.

Of course the deflationary scenario just described would involve high transition costs. So suppose that, instead of tolerating deflation, the Fed accompanied its new 20 percent reserve requirement with a plan to expand its balance sheet just enough to allow banks to meet the new requirement without having to shrink their own balance sheets. Although the new plan would avoid major changes in nominal magnitudes apart from a
substantial increase in nominal bank reserves, it would, according to standard quantity-theory reasoning, result in the same long-run real outcomes as the one just described. That is, it would lead to a new steady-state that was just as financially repressive as the one to which the deflationary alternative led.

The Fed’s actual interest on reserves policy resembles this last alternative, the difference being its use of subsidized interest on reserve payments instead of high mandatory reserve requirements to dramatically boost banks’ demand for reserves. That difference makes the interest on reserves alternative less financially repressive in so far as, instead of exposing banks to a larger required reserve tax burden, it lowers that burden, allowing banks in turn to raise their deposit rates somewhat, instead of having to reduce them as they must when they’re compelled to hold more non-interest-bearing reserves. Interest on reserves therefore has both “income” and “substitution” effects on financial intermediation, where the first increases its overall extent, while the second directs a larger share away from private borrowers and toward the Fed. An increased interest rate on both required and excess reserves will be financially repressive so long as the income effect is small relative to the substitution effect. Since the increase in banks’ net interest income consists mainly if not solely of income earned on their (relatively modest) required reserve balances, and most studies suggest that the demand for bank deposits is not very interest elastic, that’s almost certainly the case.

X. c. Reserve Hoarding and the Productivity Slowdown

One of the most disconcerting features of the post-crisis recovery has been the “great productivity slowdown” that has accompanied it (Kravis 2017). Since the start of the

---

44 For a formal demonstration of this see Dutkowsky and VanHoose (2017) who observe (ibid, p. 6): “On the one hand, an increase in the interest rate on reserves induces banks to substitute out of retail loans into excess reserves, reflected in the negative term. On the other hand, the rise in the interest rate on reserves increases revenue from required reserves. This prompts banks to demand more deposit funds and thereby allocate more of these funds to retail loans, as described by the positive term. Note that when the [required] reserve ratio equals zero, the positive term in the expression disappears. In that case, substitution, like the effect of the federal funds rate on retail loans, prevails.”

45 See, for example, Amal and Hannan (1999).
recession in late 2007, labor productivity has grown at an average annual rate of just 1.1 percent—far below the 2.3 percent average growth rate between 1947 and 2007. Many reasons have been offered for the slowdown, including a deficient supply of bank credit. As the abstract to a recent IMF study (Duval, Hong, and Timmer 2017) puts it, “the combination of pre-existing firm-level financial fragilities and tightening credit conditions made an important contribution to the post-crisis productivity slowdown.”

Such findings are not all that surprising in light of the understanding that many central banks, and the Fed especially, have embraced novel monetary policy frameworks that are highly financially repressive, and also in light of the vast theoretical and empirical literature linking such financially-repressive policies to economic underdevelopment.46 As Robert Barro (2016, p. 6) has observed regarding the U.S. case, “The dramatic rise in high-powered money was good for the Fed’s profits (most of which went to the U.S. Treasury). However, none of this was likely to contribute to productivity growth.”

On the contrary: notwithstanding then-Governor Jerome Powell’s (2017) claim that “monetary policy ... cannot reliably affect the long-run sustainable level of the economy’s growth,” by dramatically increasing the size of its credit-intermediation footprint, the Fed has almost certainly influenced U.S. economic growth—adversely. Determining the approximate size of this adverse contribution is a far from simple endeavor, in part because it means untangling the contribution of the Fed’s re-direction of credit from that of other, including regulatory “credit frictions.” The necessary research has yet to be undertaken. But that ought not to prevent policymakers from anticipating its likely outcome.47

46 See, in particular, Roubini and Sala-i-Martin (1992), De Gregorio and Guidotti (1995), and Levine (1997).
47 Silvia Merler (2018) surveys important recent research, mostly pertaining to countries other than the U.S., concerning “The Financial Side of the Productivity Slowdown.”
XI. Fiscal Consequences of the Floor System

A floor system differs from any sort of corridor systems, including the Fed's pre-crisis system, not only in its effects on the workings of monetary policy in the strict sense, or on credit allocation, but also in its broader fiscal implications. In particular, it differs in three respects. First, other things equal, a floor system results in lower Fed remittances to the Treasury than a more traditional operating framework. Second, whereas under its traditional operating system, the Fed never risked becoming insolvent, it does face such a risk, albeit perhaps a very slight one, under a floor system. Finally, whereas under its old framework the size of the Fed’s balance sheet was a crucial determinant of its monetary policy stance, that’s no longer so today. Instead, the size of the Fed’s balance sheet must be decided by other considerations.

XI. a. Treasury Remittances

The Federal Reserve’s capacity to generate interest revenues in excess of its interest expense, and to do so even despite holding safe, short-term Treasury securities, rests upon its ability to pay below-market interest rates on its liabilities. Until the crisis, the Federal Reserve’s liabilities, including both its circulating notes and its deposit balances, were non-interest-bearing. Consequently whatever interest the Fed earns on its own asset holdings is interest free and clear of any associated interest expenses; and since it earns tens of billions of dollars in net interest income, while its operating expenses amount to several billion dollars only, the Fed faces practically no risk of becoming insolvent. \(^{48}\)

Whatever income the Fed earns beyond its operating expenses is first used to pay its statutory dividends and to fund its capital surplus. The rest is remitted to the Treasury. According to a recent GAO report (United States, Government Accountability Office 2017, p. 10), since the Fed was first established the Treasury has received about 95 percent of its net earnings.

---

\(^{48}\) The Fed also earns modest amounts of non-interest revenue from fees it charges depository institutions for various services. But under the rules of the 1980 Depository Institutions Deregulation and Monetary Control Act, those fees are only supposed to just cover the Fed’s costs of supplying those services.
Other things being equal, a floor system reduces the Fed’s remittances to the Treasury by the amount of interest the Fed pays on banks’ reserve balances. Moreover, assuming that the Fed holds only short-term Treasury securities, a floor system remains more costly regardless of the quantity of excess reserves outstanding, and any corresponding change in the size of the Fed’s balance sheet. The last conclusion follows from the fact that, in an orthodox floor system, the return on bank reserves must be at least equal to that on other short-term, riskless assets, so that banks’ face no opportunity cost of holding excess reserves. In other words, the “Friedman rule” must be satisfied.

Because bank reserves are no longer a source of net income under a floor system, it’s at least conceivable that a central bank operating such a system might become insolvent—it’s conceivable, in other words, that the discounted present value of its net earnings could fall below zero. For example, suppose that the liabilities of a central bank operating with a floor system consist exclusively of bank reserve balances, and that it has no other sources of net revenue. In that case, a sufficient condition for the central bank to be insolvent is that its operating expenses are greater than zero! Any obligation to pay dividends or contribute to a surplus fund might also suffice.49

In practice, though, as can be seen in Figure XI.1, the Fed’s floor system, far from rendering it insolvent, has allowed it to earn and remit to the Treasury more revenue than ever. There are three reasons for this. First, while the Fed now pays interest on banks’ reserve balances, and also (through its ON-RRP program) on balances held by GSEs, it still doesn’t pay any interest on its other liabilities, including outstanding Federal Reserve

---

49 The Federal Reserve System’s annual operating expenses for 2016 were $6,676 million. Until recently all Fed member banks received dividends equal to 6 percent of their paid-in capital. However, the Fixing America’s Surface Transportation (“FAST”) Act of 2015 reduced the dividend for banks with over $10 billion in assets to the lesser of the 10-year Treasury note rate or 6 percent. The FAST Act also placed a $10 billion cap on the Fed’s surplus account.

In claiming that it is “perfectly feasible” for a central bank to pay interest on reserves according to the Friedman-rule policy “since the interest it must pay on its outstanding reserves is covered by the interest it earns on the assets it purchased when those reserves were created,” Keister, Martin, and McAndrews (2015, p. 4) overlook central banks’ non-interest expenses and obligations. They are right, on the other hand, in faulting Berentsen et al. (2014) for exaggerating the prospects for insolvency under a floor system by assuming, in their baseline analysis at least, that the central bank has liabilities, but no assets!
notes. Those notes therefore “effectively provide an interest-free loan to the Federal Reserve, which, given Federal Reserve remittance policies, implies they represent an interest-free loan to the Treasury” (Wall 2015).

Although currency now makes up only about one-third of the Fed’s total liabilities, while bank reserves make up 60 percent, whereas before the crisis the respective figures were about 90 percent and two percent, the Fed’s net earnings depend, not on currency’s share, but on its total amount, which has almost doubled since August 2008.

A doubling of the outstanding stock of currency cannot, however, alone account for a near tripling of the Fed’s Treasury remittances, and especially so considering the general post-crisis decline in interest rates, including yields on short-term Treasury securities. Nor can the vast post-2008 increase in banks’ excess reserves alone account for it, for as we’ve seen, that growth alone would normally cause the Fed’s interest expenses to grow in step with its interest earnings. Instead the difference is due to the Fed’s having purchased long-term securities, which allowed it to profit from the prevailing term premium, though only by taking on considerable interest-rate risk.

Indeed, if there is any serious risk that the Fed’s floor system might render it insolvent, it arises from the possibility that, in consequence of its attempts to profit in the short-run from increased bank excess reserve holdings by assuming more duration risk, it suffers losses large enough to wipe-out the profits it earns on its non-interest-bearing
liabilities, in which case it would have no choice but to either be recapitalized by the Treasury or to abandon its floor system, with its implicit Friedman rule, and resort to higher inflation as a means for increasing its monopoly profits (Bassetto and Messer 2013, p. 413; see also Del Negro and Sims, 2015).

According to projections by Fed economists and others (e.g., Carpenter et al. 2013; Greenlaw et al. 2013; Del Negro and Sims 2015; Ferris et al. 2017; and Cavallo et al. 2018), the Fed’s losses due to rising interest rates are likely to be modest compared to the present value of its future monopoly profits. Still, as Figure XI.2 shows, those losses are likely to be far from trivial. It’s owing to them, and not to the unwinding of its balance sheet, that the Fed expects its Treasury remittances to decline substantially in the coming years (Figure XI.3).50

Figures XI.2 and XI.3: Projected Fed Losses (left) and Treasury Remittances (right) (Source: Reproduced from Ferris et al., 2017)

50 Fed earnings projections are, furthermore, sensitive to assumptions made about future Treasury security term premiums. Cavallo et al.’s claim that a larger Fed balance sheet will ultimately mean increased Fed remittances to the Treasury, for example, rests on the assumption that term premiums will eventually rise to their historically positive values. If one assumes instead that the term premium, which has been trending downward for many years and is presently negative, will remain so in the future, the larger the Fed’s balance sheet, the lower its Treasury remittances will be. That case of course amounts to a further argument for keeping the Fed as small as possible. (I thank Bill Nelson, Chief Economist and Head of Research at The Clearing House Association, for drawing my attention to this issue.)
XI. b. Other Fiscal Consequences of a Floor System

To summarize: although the Fed’s floor system is unlikely to render it insolvent anytime soon, the system is bound to have a higher fiscal cost than any sort of corridor system, because it generally calls for a higher interest rate on reserve balances, including balances that banks would hold even without the encouragement of interest on reserves, and a corresponding reduction in the Fed’s Treasury remittances.

From a strictly utilitarian perspective, this increased fiscal cost is not entirely a bad thing. For although it reduces the Fed’s surplus, and with it the U.S. Treasury’s revenue, that reduction in producer’s surplus has as its counterpart an increase in consumers’ surplus, consisting of increased commercial bank earnings at least some of which are passed on to commercial bank depositors in the form of higher rates on commercial bank deposits. The Friedman rule is, after all, supposed to lead to precisely this outcome. However, for reasons discussed earlier, because a floor system generally involves interest payments in excess of those required to implement the Friedman rule, a corridor system with a below-market IOER rate is better suited for achieving an “optimum” return on reserve balances.

Furthermore, as David Beckworth (2017) has noted, such a strictly utilitarian perspective ignores “an important political-economy consideration,” to wit: that the main recipients of interest on reserves are some foreign banks and the very largest U.S. banks. This, Beckworth observes, makes for some particularly “bad optics”:

Think about the implications: the banks that were bailed out during the crisis and the banks owned by foreigners are getting most of the IOER payment. This is a perfect storm of financial villains for both the political left and the right... [T]he bad optics will only look worse if the Fed’s balance sheet does not shrink as interest rates go up. For the IOER payment will go up too.
For this reason, as Cavallo et al. (2018, p. 14) also recognize, “large payments to the banking system could result in political pressure against maintaining or raising the IOER.”

But by far the most serious fiscal danger posed by a floor system stems from the fact that, under such a system, the size of the Fed’s balance sheet ceases to have a crucial bearing upon the stance of monetary policy. Although—as I’ve noted above—the much-vaunted monetary control advantages of this “decoupling” of balance sheet policy from monetary policy are quite dubious, it has a very real political disadvantage, to which both Charles Plosser (2017) and I (Selgin 2017b) have drawn attention. As Plosser (2017, p. 6) explains, the fact that, under the floor system, the Fed’s balance sheet can be made arbitrarily large “without impacting the conduct of monetary policy” (ibid, p.6), raises a crucial question, namely, “How and who will determine the amount of excess reserves created”? (ibid, p. 7):

Making the Fed’s balance sheet unrelated to monetary policy, opens the door for the Fed to use its balance sheet for other purposes. For example, the Fed would be free to engage in credit policy through the management of its assets while not impinging on monetary policy. Indeed, the Fed’s balance sheet could serve as a huge intermediary and supplier of taxpayer subsidies to selected parties through credit allocation (ibid.)

Because it makes the size of the Fed’s balance sheet “a free parameter” (ibid, p. 8), the new arrangement “also opens the door for Congress (or the Fed) to use the balance sheet for its own purposes (ibid, p. 7). “Congress,” Plosser observes,

---

51 For examples of these “bad optics” at work in popular publications see Durden (2017) (ZeroHedge), “Is the Federal Reserve Giving Banks a $12bn Subsidy?” (2017) (The Economist), Richter (2018) (Business Insider), and Heller (2016) (American Banker). The last asks his readers to “Think of the potential political firestorms that will erupt as commercial banks are paid an amount equivalent to $100 billion per year that could otherwise be used to shore up federal resources.”
would be free to lobby the Fed through political pressure or legislation to manage
the portfolio for political ends. Imagine a Congress proposing a new infrastructure
bill where the Fed was expected, or even required, to buy designated development
bonds to support and fund the initiative so taxes could be deferred. This would be
very tempting for Congress. Indeed, in testimony before Congress I was asked why
the Fed shouldn’t contribute “its fair share” to an infrastructure initiative. Image
the lobbying for the Fed to purchase “build America bonds” issued by the Treasury
to fund infrastructure initiatives. … Of course, this just represents off-budget fiscal
policy… Congress will undoubtedly find many “appropriate” uses for the Fed’s
balance sheet and could do so and claim it doesn’t interfere with the independence
of monetary policy” (ibid, p. 8).

One may also view the problem as one stemming from the fact that in a floor
system even very large Fed asset purchases no longer result in substantial changes in the
price level. “Fear of inflation,” I observed not long before Plosser’s essay appeared, “has
kept a lid on the Fed’s financing of deficits” for decades. But in the new set-up, that
constraint no longer operates:

Now, if the Fed decides to gobble-up still more Treasury or government-agency
securities, putting a like sum of fresh reserves at banks’ disposal, it can still keep
inflation at bay by hiking the IOER rate enough to bribe banks to hoard the
reserves instead of lending them out (Selgin 2017b).

Nor can growth in the Fed’s balance sheet itself be counted upon to provoke a public
outcry:

While inflation makes headlines, reserve hoarding doesn’t. That’s why the Brave
New World of interest on reserves is so dangerous. Faced with the usual pressure
to help the government pay its bills, Fed officials, and a pliant or weak Fed Chair
especially, might cave-in to the government’s demands while still meeting the
Fed’s inflation targets. In theory they could go on meeting the government’s
demands until every penny in bank deposits is financing some government spending, leaving nothing for private-sector borrowers.

Here, as in Plosser’s argument, the concern is that, because it “creates the opportunity and incentive for political actors to exploit the Fed’s balance sheet to conduct off-budget fiscal policy and credit allocation,” the Fed’s floor system could “undermine Fed independence and politicize the Fed to a far greater degree than currently exists” (Plosser 2017, pp. 13-14). There is even reason to fear that, by exploiting the Fed’s floor system in pursuit of its own projects, Congress might ultimately cause the Fed to go broke after all, thereby dealing a further, and very severe, blow to its independence.
XII. The Floor System and Policy Normalization

XII. a. The Fed’s Plan

Ever since the Fed began its large-scale asset purchases, Fed officials have been promising that, once recovery from the crisis was complete, they would begin a process of monetary policy “normalization.” In particular, they promised to eventually reduce the size of the Fed’s balance sheet, though they only announced a specific plan for doing so relatively recently. They’ve also promised to get the Fed’s policy rate settings back to “normal” levels.

Regarding the latter aim, most Fed officials now take for granted a long-run “normal” fed funds rate level of about 3 percent, reflecting an assumed normal real rate—“r-star,” in Fedspeak—of one percent, plus the Fed’s 2 percent inflation target. Having the effective fed funds rate approach 3 percent is therefore also part of their normalization strategy. To judge by FOMC members’ most recent projections, as shown in Figure XII.1, that goal, or something close, will be reached within the next several years, which is to say, while the Fed is also in the process of shrinking its balance sheet.

Figure XII.1: FOMC R-Star “Dot Plot,” September 2017
(Source: FOMC Projection Materials, Sept. 20, 2017)
The Fed’s plan for shrinking its balance sheet, which it unveiled in June 2014 and started to implement in October 2017, calls for it to shed $1.5 trillion assets between now and 2022, bringing its balance sheet to $3 trillion, or to about 15 percent of projected 2022 GDP (Board of Governors 2017a). James Hamilton (2017) has illustrated the progress of the Fed’s unwind in several nice charts, one of which is reproduced as Figure XII.2 below. Besides showing the slow pace of the Fed’s unwind, the chart also shows that once the balance sheet declines to $3 trillion or so it is expected to start swelling again.

![Figure XII.2: Actual and Projected Federal Reserve Assets, 2002-2025](Source: Hamilton 2017)

The Fed’s normalization plan has been criticized on several grounds. First, because the Fed’s unwind allows for no outright asset sales, while placing varying limits or “caps” on the value of assets it will allow to roll-off its balance sheet during any particular month, its pace is disappointingly languid. Consequently the Fed’s credit footprint, and its influence on credit allocation, will also decline only very slowly. The Fed might easily have arranged to unwind at a faster yet still regular pace by combining higher roll-off caps with occasional active asset sales during low roll-off periods. According to Willem Buiter (2010), the impediments to a quick unwind are not technical but political; they consist of

---

51 Having begun last October with maximum roll-offs of $6 billion in Treasuries and $4 billion in MBS, the Fed plans to increase the caps by those same amounts every three months for a year, after which they’ll remain fixed at $30 billion for Treasury securities and $20 billion for MBS. Whenever the value of maturing assets of either sort exceeds its assigned cap, the Fed plans to reinvest the difference.
resistance from Treasury authorities seeking to avoid a sharp decline in the value of their securities as well as the Fed’s own reluctance to draw attention to

the true extent of the central bank’s quasi-fiscal activities during the crisis and its aftermath.” The large-scale ex-ante and ex-post quasi-fiscal subsidies handed out by the Fed and to a lesser extent by the other leading central banks, and the sheer magnitude of the redistribution of wealth and income among private agents that the central banks have engaged in could (and in my view should) cause a political storm. Delay in the dropping of the veil is therefore likely (ibid., p. 23).

Second, even at its minimum level, the Fed’s balance sheet will be far larger, relative to the size of the U.S. economy as a whole, than it was prior to the crisis: were the Fed intent on getting its balance sheet back on its pre-crisis trajectory by the end of 2020, it would be planning on slimming down by at least another $500 billion, and perhaps by another $1 trillion. As Ulrich Bindseil (2016, p. 227) explains, there are good reasons for wishing to see the Fed pursue this more aggressive diet:

While in crisis times, there are a number of justifications for lengthening the central bank balance sheet through LSAPS or LOLR operations … a lean balance [in normal times] suggests that the central bank focuses on the core of its mandate. Moreover, a lean balance sheet is a sign of well-functioning financial markets and a healthy economy because the central bank is neither used as intermediary by the banking system, nor does the central bank see a need to engage in special crisis measures such as LSAPs.

“An outstandingly lean central bank balance sheet,” Bindseil goes on to say, was among the Fed’s principal merits before the crisis, when its total assets were “only around 1.1 times the total amount of bank notes in circulation” (ibid, p. 228).

Finally, the Fed’s determination to raise its policy rate (or rates) to a preconceived “normal” level, and to do so within a relatively rapid span of time, seems imprudent, and may, if pursued obstinately, ultimately cause it to further delay its planned balance sheet
reduction, or even to abandon it altogether. The FOMC, Yellen said in her July 2017 testimony, was

prepared to resume reinvestments if a material deterioration in the economic outlook were to warrant a sizable reduction in the federal funds rate. More generally, the committee would be prepared to use its full range of tools, including altering the size and composition of its balance sheet, if future economic conditions were to warrant a more accommodative monetary policy than can be achieved solely by reducing the federal funds rate (Yellen 2017b).

XII. b. A New Normal

But to harp on these problems is to overlook an elephant in the room, to wit: the Fed’s floor system, and its desire to keep it. Although the Fed never mentions that desire in laying-out its normalization plan, its presence supplies a parsimonious explanation for several of that plan’s details, and most obviously for the limited extent of the Fed’s planned diet. Thus Hamilton (2017) argues, correctly, that so long as the Fed sticks to its present operating framework, the volume of reverse repos it conducts, as well as the value of Treasury balances held with it, will remain highly volatile, and that it will take "a lot of reserves sloshing around the system to cover that kind of variation." However, as Hamilton also notes, it would be easy enough for the Fed to make all that extra cash unnecessary if it wanted to, the most straightforward solution being that of "moving to a true corridor system for controlling interest rates." Because a more abundant supply of

53 In fact there is considerable disagreement, even within the Fed, concerning the likelihood that r-star (the equilibrium real federal funds rate) will reach one percent within the next several years. For example, in their recent San Francisco Fed study Jens Christensen and Glenn Rudebusch (2017b) use Treasury Inflation-Protected Securities (TIPS) prices to arrive at what they consider to be especially reliable r-star estimates and projections. They conclude that as of December 2016 r-star stood close to zero, and that it “is more likely than not to remain near its current low for the foreseeable future” (ibid., p. 27). Elsewhere the same authors (Christensen and Rudebusch 2017a) observe that “For policymakers and researchers, the equilibrium interest rate provides a neutral benchmark to calibrate the stance of monetary policy: Monetary policy is expansionary if the short-term real interest rate lies below the equilibrium rate and contractionary if it lies above. Therefore, determining a good estimate of the equilibrium real rate has been at the center of recent policy debates.”
bank reserves is a necessary feature of any floor-type operating system, the Fed’s inclination to retain such a system may at least partly account for its desire to stay pudgy.\footnote{Other arguments have, to be sure, been advanced for having the Fed retain a much-enlarged balance sheet. For these see Greenwood (2016) and Buiter, Jensen, and Rojas (2017). For an able reply, see Goodhart (2017). Among other points, Goodhart observes that while these authors “present arguments for not shrinking a Central Bank’s [specifically, the Fed’s] balance sheet from its present level, they advance none for further increasing it. Since the present level was largely attained by happenstance, proposals to maintain that (ad hoc) level suggest a lack of determining principles, and, perhaps a certain doubt” concerning the “strength of the underlying arguments.”}

The Fed’s determination to stick to a floor system marks a major switch from its original understanding of what policy normalization entailed. Back in February 2010, when he first testified before Congress on what was then still referred to as the Fed’s “exit strategy,” Ben Bernanke told Congress that the Fed “anticipates that it will eventually return to an operating framework with much lower reserve balances than at present and with the federal funds rate as the operating target for policy” (Bernanke 2010a). In a footnote to his written testimony, Bernanke (ibid.) made it clear that he had a corridor system in mind:

> The authority to pay interest on reserves is likely to be an important component of the future operating framework for monetary policy. For example, one approach is for the Federal Reserve to bracket its target for the federal funds rate with the discount rate above and the interest rate on excess reserves below. Under this so-called corridor system, the ability of banks to borrow at the discount rate would tend to limit upward spikes in the federal funds rate, and the ability of banks to earn interest at the excess reserves rate would tend to contain downward movements (ibid, n9).

Although Bernanke adds, in the same note, that “other approaches are also possible,” and that the Fed “has ample time to consider the best long-run framework for policy implementation,” the Fed was evidently inclined to return to an arrangement differing only modestly from its pre-crisis system. Apart from being reasonably consistent
with a literal understanding of “normalization,” that plan would have realized, as a floor system could not, Bernanke’s hope that the Fed would eventually settle on an operating framework that would not “impose costs and distortions on the banking system.”

That top Fed officials have since had a change of heart is evident, not only from the details of the Fed’s normalization plan, including a long-run balance sheet roughly three times larger, relative to GDP, than its pre-crisis counterpart, but from some of their recent statements. For example, although Janet Yellen (2017a), in her June 2017 remarks to the press, observed then that “the longer run normal level of reserve balances will ... depend on the committee’s eventual decisions about how to implement monetary policy most efficiently and effectively in the longer run,” and that the Fed still had plenty of time left to decide on its eventual operating framework, she went on to say that our current system is working well and has some important advantages. In particular, it’s simple and efficient to operate, does not require active management of the supply of reserves, and, most importantly, provides good control over the federal funds rate and effective transmission of changes in the federal funds rate to broader money market rates. And because our current system is likely compatible with the much smaller quantity of reserves, our plan for gradually reducing our balance sheet does not constrain the Committee’s future options for how to implement monetary policy (ibid, pp. 5-6).

Yellen observed, furthermore, that “changing the target range for the federal funds rate” would remain the Fed’s “primary means of adjusting the stance of monetary policy” and that the Fed did not intend to treat the Fed’s balance sheet as “an active tool for monetary policy in normal times” (my emphasis). Taking that last remark to rule out, not just Quantitative Easing, but also ordinary open-market operations, as means for influencing the fed funds rate, Yellen’s statement suggested that the Fed was inclined to stick to a floor system.
In a speech he gave two weeks earlier, at the Economic Club of New York (Powell, 2017), then-Governor Powell was forthright concerning his own preferences:

[When the [Federal Open Market] Committee discussed using a floor system as part of its longer-run framework, I was among those who saw such an approach as "likely to be relatively simple and efficient to administer, relatively straightforward to communicate, and effective in enabling interest rate control across a wide range of circumstances."

Some have advocated a return to a framework similar to the pre-2007 system, in which the volume of reserves would likely be far below its present level and the federal funds rate would be managed by frequent open market operations. This "corridor" framework remains a feasible option, although, in my view, it may be less robust over time than a floor system.

So why not stick with the floor system? First, because, as we’ve seen and despite what Yellen and Powell say, it has never worked well. True, it has succeeded in keeping the effective fed funds rate within the Fed’s “target range.” But in a system in which fed funds activity is almost all devoted to arbitraging the difference between the upper and lower limits of that range, such success is nugatory. The Fed’s new framework has also succeeded in the sense that changes in the IOER rate have led to like changes in other short term rates. But as has been noted, these achievements refer to the Fed’s intermediate policy objectives only, rather than to its ultimate policy goals. Assessed in light of those ultimate goals, the Fed’s new operating framework can only be judged a failure.

The inherent drawbacks of floor systems, combined with the challenge of unwinding the vast balance sheet increases that such systems instigated, may explain why the Fed, the Bank of England, and the ECB alone still rely on floor arrangements. Other central banks that entered the crisis with corridor systems of some kind either (1) stuck to those systems all along (e.g. the Riksbank and the Reserve Bank of Australia); (2)
switched to a floor system temporarily, and then reverted to a corridor (the Bank of Canada); or (3) switched first to floor systems and then to “tiered” or “quota” systems in which reserves can earn the policy rate only up to a certain limit, after which they earn lower, if not negative, rates (the Swiss National Bank, Bank of Japan, Reserve Bank of New Zealand, and Norges Bank). Significantly, at least two of the central banks that switched from floor systems to tiered ones—those of New Zealand and Norway—did so in order to counter “banks’ insatiable appetite for reserves,” and thereby revive interbank lending (Sellin and Åsberg 2014; see also Berhardsen and Kloster 2010).55

XII. c. A Plan for Genuine Normalization

If the Fed’s “normalization” plan has been informed by Fed officials’ desire to keep the present floor system in place, moving from that system to a corridor system, like the one Fed officials had in mind in 2006 (when they first gained permission to pay interest on bank reserves), calls for a very different, more ambitious, and more challenging plan.

The present system rests on two requirements. First, it requires that excess reserves yield more interest than other short-term securities. Second, it requires that supply of reserves (federal funds) be more than sufficient to equate quantity supplied and quantity demanded at the going IOER rate. A corridor system requires, on the other hand, that banks incur an opportunity cost by holding excess reserves, so that they will be inclined to economize on reserves, and to make use of overnight interbank loans to meet occasional, short-term liquidity needs. The last condition guarantees that there will be an active federal funds market, with a market-clearing fed funds rate that can serve as an intermediate policy target, controllable by means of open-market operations.

55 According to Norges Bank (2011), its switch from a floor to a quote system had two objectives. “One was to stop growth in central bank reserves. The other was to generate more interbank activity in the overnight market as rising quantities of reserves in the banking system entailed a risk that Norges Bank would assume some of the functions of the interbank market. The role of the central bank is to steer the total amount of reserves, while banks’ should redistribute the reserves overnight in the interbank market.”

Birdsell (2016) undertakes a thorough assessment of alternative central bank operating frameworks, the conclusion of which is that “the symmetric corridor continues to perform well” relative to other alternatives.
Moving from a floor system to a corridor is challenging because the move will entail an ultimately-desirable revival of the once-moribund base money multiplier. As we’ve seen, for M2 that multiplier is now less than half what it was before September 2008, so that reviving it could result in a doubling, or more, of the equilibrium money stock, nominal income, and the price level, other things equal. To avoid missing its policy targets, the Fed would have to counter the multiplier’s revival either by resorting to offsetting asset sales or by otherwise immobilizing excess reserves that banks are no longer inclined to hold. Given the huge quantity of excess reserves now outstanding, the Fed could find itself having to sell a corresponding volume of assets in a hurry, upsetting markets. “No one,” Karen Petrou (2016) has observed, “can forecast how still more trillions of assets set loose would fare, but it’s clear that, at the very least, market volatility would spike and financial stability would be endangered.”

An obvious step the Fed can take to limit both the adverse effects of dealing adequately with a multiplier revival, and the harm that would ensue from not dealing with it adequately, is to shrink its balance sheet to the absolute minimum level required to maintain a floor system before it “normalizes” its IOER rate, ultimately by allowing it to fall below comparable market rates. In that way, both the absolute extent of the monetary expansion a multiplier revival would entail, and that of further asset sales needed to prevent that expansion, would also be minimized.

And the Fed might shrink its balance sheet, while still preserving its floor system, to a far greater extent than its present normalization plan would allow. In defense of the Fed’s plan, Ben Bernanke (2017) observes that by December 2008, when the Fed’s floor system was firmly established, “bank reserves were about $800 billion,” and that “taking into account growth in nominal GDP and bank liabilities” suggests that “the critical level of bank reserves needed to implement monetary policy through a floor system seems likely to be well over $1 trillion today, and growing.” However, the fact that banks held $800 billion in reserves in December 2008 doesn’t at all imply that a smaller quantity of reserves would not have met the requirements of a floor system equally well. First of all,
although it wasn’t until December 2008 that the Fed officially switched from a single-value fed funds rate target to an IOER-based target range, the Fed was already operating a de-facto floor system by the end of October, when reserves were much lower. Second, as Stephen Williamson (2017a) points out, Canada’s brief floor-system experience shows that, although a floor system allows central banks to vastly expand the stock of bank reserves without significantly altering the stance of monetary policy, such a system doesn’t actually require all that many more reserves than would suffice for a corridor system:

How much [sic] reserves did it take to make [Canada’s] floor system work? The Bank targeted overnight reserves to $3 billion (Canadian) over this period. To get an idea of the order of magnitude, a rule of thumb is that the Canadian economy is roughly a multiple of 10 of the US economy, so this quantity of reserves is roughly comparable to $30 billion in the US. We need to account for the fact that there are reserve requirements in the US, and none in Canada, and that the US institutional setup is very different (many more banks for example). But, I think it’s hard to look at the Canadian experience and think that it takes as much as $1 trillion in interest bearing Fed liabilities to make a floor system work in the US, as Bernanke is suggesting. I would be surprised if we needed as much as $100 billion.

In other words, allowing for the demand for currency, the Fed might operate its floor system today with a balance sheet of less than $1.7 trillion dollars, and might do so in 2020 with a balance sheet of below $2 trillion, or over $1 trillion less than that provided for by its current normalization plan. Indeed, since that plan, in extrapolating from recent trends, may well overstate the future demand for currency, a 2020 balance sheet of considerably less than $2 trillion might suffice.56 Nor are these estimates especially conservative. Ricardo Reis (2016, p. 463), for one, goes further, maintaining that “keeping the market for reserves saturated is consistent with returning to a lean central-bank

56 As Hugh McCulloch observes (2018, p. 7), it has been mainly owing to the persistence of very low interest rates that “currency in circulation has almost doubled since 2007, while the nominal economy has only grown about 33%,” and that “this situation cannot be expected to continue forever.”

Nor need the Fed take much longer than until 2020 to achieve a substantially more aggressive unwind. As I’ve noted, its present plan, with its exclusive reliance upon capped roll-offs of maturing assets, is extremely conservative, whereas by relaxing the roll-off caps, while undertaking outright sales to compensate for low roll-off periods, the Fed could speed-up its unwind considerably while still maintaining a smooth unwind schedule.

Until the Fed’s balance sheet is reduced to its minimal level consistent with maintaining its floor system, the Fed should continue to implement policy by adjusting its IOER and ON-RRP rates as needed to achieve its policy objectives—and not merely according to some preconceived notion of a “normal” fed funds rate and announced schedule for arriving at it. But once the minimum reserve level is reached, it can proceed to the final steps toward genuine policy normalization. These steps are the most difficult, because they must be carefully coordinated. One consists of a gradual reduction in the IOER rate aimed at raising the opportunity cost to banks of holding excess reserves sufficiently to encourage banks to once again dispose of those reserves; the other consists of a further reduction in the nominal quantity of bank reserves, aimed at encouraging banks to routinely rely once again upon interbank borrowings to meet their liquidity needs. Ideally, because the one step reduces the demand for reserves, while the other reduces their supply, the steps should be policy neutral.\textsuperscript{57} However, they will ultimately

\textsuperscript{57} As Larry Wall (2017) observes, the balancing act won’t necessarily be easy: “If the amount of sales ordered by the FOMC was too small, the resulting excessive stimulation would likely result in higher inflation. If the amount of sales was too large, the resulting excessive tightness could cause the economy to go into a recession.” Wall is, of course, correct. But the problem he describes is one common to any scheme that relies upon balance sheet adjustments, routine or otherwise, to achieve a desired monetary policy stance, during an episode of volatile reserve demand. There is some risk, on the other hand, that even starting with a minimized floor-system balance sheet, the asset sales needed to compensate for a sudden revival of
change the way policy is implemented, by serving to revive the federal funds market, so that the Fed will be able to abandon its administered rate target “range” and to once again target the effective fed funds rate.

During the transition itself, however, the Fed will need some way to both gauge and communicate its policy stance effectively. The challenge here is one Ben Bernanke (2010a, p. 10) addressed when he outlined the Fed’s original “exit” strategy back in February 2010. To meet it, he said, the Fed might temporarily switch to “communicating the stance of policy in terms of another operating target, such as an alternative short-term interest rate.” Fortunately, several very good alternatives are now readily available, consisting of benchmark rates developed recently by the Federal Reserve Bank of New York, based on private Treasury-secured overnight repos (Federal Reserve Bank of New York 2017).

Once the federal funds market is functioning as it did before the crisis, with substantial volumes of interbank lending and borrowing, the Fed can return to its traditional reliance upon open-market operations to keep the effective federal funds rate at its chosen target level. Allowing that the Fed still pays interest on bank reserves, albeit at a rate that remains below the policy target rate, the new arrangement will differ only slightly from the Fed’s pre-October 2008 system, the difference being that between the Fed’s old system and an orthodox corridor system of the sort Fed officials had in mind in 2006, when they were first granted permission to pay interest on bank reserves.

---

wholesale bank lending might be large enough to disturb the markets for the assets being sold. Fortunately, the Fed has a ready-made solution to this problem, in the shape of the Term Deposit Facility it established in early 2017 “to facilitate the conduct of monetary policy by providing a tool that may be used to manage the aggregate quantity of reserve balances held by depository institutions” (Central Bank Central 2017). The term deposit auctions undertaken by that facility serve, like Fed asset sales, to drain reserves from the banking system for the term of the auctioned deposits, but do so without disrupting asset markets. The Fed might therefore rely upon such auctions, as it passes from a floor to a corridor regime, to achieve a more gradual, further reduction of its balance sheet.
Achieving a genuine normalization of monetary policy, including a non-disruptive return to a corridor-type operating system, is, I’ve tried to suggest, difficult, but perfectly possible. And even if it would be considerably more difficult than the Fed’s present plan, which would keep the floor system in place indefinitely, that greater difficulty would not be a good reason for sticking to the current plan. As I’ve tried to show in some detail in these pages, the Fed’s floor system of monetary control is both unreliable and inefficient. Because it severs the link between reserve creation and monetary expansion, it makes achieving monetary stimulus, even by means of extraordinary asset purchases, extremely difficult; and because it has the Fed borrowing heavily from private intermediaries, it replaces private-sector lending with lending to the U.S. Treasury and other government agencies. Finally, the Fed’s floor system is inconsistent with the intent of the 2006 law that first granted the Fed the right to pay interest on bank reserves.

Ideally, Fed officials can still be convinced, in light of the floor system’s serious shortcomings, to alter their plans so as to eventually switch to an orthodox corridor system. But whether they can or can’t, Congress should uphold the spirit of the 2006 law allowing the Fed to pay interest on reserves, either by clarifying that law’s stipulation that rates paid on reserves not “exceed the general level of short-term interest rates,” 58 or by limiting interest payments to banks’ required reserve balances, while allowing the Fed a period of no more than two years to comply with the revised statute. Should Fed officials still wish to have a floor system, they can always seek Congress’s permission, by proposing new legislation specifically granting them the necessary powers, which they’ve exercised thus far only by flouting the law. That, after all, is how democratic government is supposed to work.

---

58 Appendix II suggests how the 2006 law might be amended to achieve this end.
Appendix I: Bank Lending and the Interest Rate on Excess Reserves

The argument of section VII.b, that the Fed’s positive IOER rate can cause a bank to lend less even when the IOER rate is substantially less than the bank’s net interest margin, agrees with the implications of many standard models of bank optimization. I offer here an example using a simplified version of Timothy Hannan’s (1991) model, which is itself based on Michael Klein’s (1971) “A Theory of the Banking Firm.” My model differs from Hannan’s in allowing for only one category of deposits, instead of several, and in assigning to reserves the part that securities play in Hannan’s model. The last change reflects the fact that the Fed’s IOER rate has generally exceeded comparable rates on relatively safe, short-term securities, so that reserves strictly dominate securities in risk-adjusted return.

A representative bank maximizes its profit on an asset portfolio consisting of reserves \((R)\) and \(N\) kinds of loans \((L)\), differing in their riskiness and maturity, among other factors, financed by deposits \((D)\):

\[
\pi = \sum_{n}^{N} \left( r_{L}^{n} - c_{L}^{n} \right) L_{n} + r_{R} R - (r_{D} + c_{D}) D,
\]

where \(r_{L}^{n}\) is the interest rate on loans of type \(n\), and \(r_{R}\) and \(r_{D}\) are the interest rates paid on reserves and deposits, respectively. \(c_{L}^{n}\) is the non-interest expense, including loan-loss provisions, associated with a loan of type \(n\), and \(c_{D}\) is the non-interest expense associated with maintaining and servicing deposits.

Banks maximize profit subject to the balance sheet constraint,

\[
\sum_{n}^{N} (L_{n}) + R = D + K,
\]

where \(K\) stands for bank capital. The first order condition for profit maximization is

\[
\frac{\partial \pi}{\partial r_{L}^{n}} = L_{n} + r_{L}^{n} \frac{L_{n}}{dr_{L}^{n}} - (r_{R} + c_{L}^{n}) \frac{dL_{n}}{dr_{L}^{n}} = 0.
\]

Defining the elasticity of demand for a type \(n\) loan as
\[ e^n_L = - \left( r^n_L \right) \left( \frac{dL_n}{dr^n_L} \right) > 0, \]

that condition can be rewritten as

\[ r^n_L = (r^n_R + c^n_L) \left[ \frac{e^n_L}{e^n_L - 1} \right], \]

which shows that the bank’s lending rate will increase as either the IOER rate or the non-interest costs of lending increases, and will decline as the elasticity of loan demand increases. To paraphrase Hannan (p. 71), “The optimal loan rate is determined by the extent to which it is profitable to replace [reserves] with loans in the bank’s portfolio.” In other words, a higher IOER rate raises loan rates by causing banks to hoard cash instead of lending it.

So long as loan demand is elastic, the equilibrium loan rate will exceed the IOER rate. In the limit, as elasticity approaches infinity (as it does in the case of a banking industry that’s either perfectly competitive or perfectly contestable), equilibrium lending rates will exceed the IOER rate by the non-interest cost of lending:

\[ r^n_L = r^n_R + c^n_L. \]

Note that, if the non-interest expense associated with any loan type is zero, then the interest rate charged on those loans will equal the IOER rate. This case is like the one illustrated in Figure VII.5, where the marginal loan rate is assumed to be equal to the IOER rate. Product differentiation, to the extent that it exists, will widen the gap between equilibrium loan rates and the IOER rate still further.

Finally, let’s consider what all this implies concerning the relation between net interest margins and the IOER rate. For the perfectly competitive (or contestable) market case, the net interest margin (NIM) for a loan of type \( n \) is

\[ NIM^n_n = r^n_L - r_D = r^n_R + c^n_L - r_D. \]

The net interest margin for a type \( n \) loan will therefore exceed the IOER rate by the difference between the non-interest and interest expenses associated with the loan. In recent years, with deposit rates at extremely low levels, if not at zero, that difference has been substantial.
Since a bank’s net loan income is equal to its net interest margin minus the non-interest expense of lending, including loan loss provisions, the equilibrium condition can be restated to be that a bank’s marginal net loan income, expressed as an annual percentage, will tend to be equal the difference between the IOER rate and the interest rate on deposits.
Appendix II: Proposed Amendment to the 2006 Financial Services Regulatory Relief Act

As written, the 2006 Financial Services Regulatory Relief Act allows the Federal Reserve to pay interest on reserves “at a rate or rates not to exceed the general level of short-term interest rates” (12 U.S.C. 461 (b)(12)(A)). However, in drafting the final rules implementing the statute, as amended by the 2008 Emergency Economic Stabilization Act (which altered the timing of the 2006 act’s provision only), Fed officials determined that for that purpose "'short-term interest rates' are rates on obligations with maturities of no more than one year, such as the primary credit rate and rates on term federal funds, term repurchase agreements, commercial paper, term Eurodollar deposits, and other similar instruments ("Regulation D: Reserve Requirements for Depository Institutions,” Federal Registrar, June 22, 2015, p. 35567)." Because the Fed’s primary credit rate is rate it administers, rather than a market-determined rate, and because that rate is always set above the interest rate paid on bank reserves, the Fed’s understanding of the meaning of “general level of short-term interest rates” allows any interest rate it chooses to pay on reserves to satisfy the statute’s requirement!

To ensure that the Federal Reserve abides by the spirit of the 2006 law, the meaning of the phrase “general level of short-term interest rates” must be clarified. The interest rates that are most suitable for that purpose are market-determined rates on instruments similar in duration and risk to the reserve balances on which the Fed is authorized to pay interest. Because reserve balances are essentially very low risk assets of zero maturity, interest rates on private overnight repurchase agreements collateralized by Treasury securities are close market-determined equivalents.

Although there is no one uniform overnight repo rate, the Federal Reserve Bank of New York has recently conducted extensive research aimed at establishing overnight repo benchmark rates, using transaction level data. Based on this research, the New York Fed has developed a “Broad Treasury financing rate” that’s perfectly suited to serve as an IOR benchmark rate, that is, as a reference “general” rate for the purpose of implementing the
In light of this, Congress should amend the 2006 Act by having “the general level of short-term interest rates” refer specifically to some value or values of the “Broad Treasury financing rate. For example, the text of the necessary law might read as follows:

“Section 19(b)(12) of the Federal Reserve Act (12 U.S.C. 461(b)(12)) is amended by inserting after Subparagraph (C)

‘(D) General level of short-term interest rates defined.—

For purposes of this paragraph, the term “general level of short-term interest rates” shall be defined as the average value over the preceding six-week interval of the Federal Reserve Bank of New York’s benchmark Broad Treasury financing rate on overnight repurchase agreements”

---

59 For details see Bayeux et al. (2017) and Federal Reserve Bank of New York (2017),
References


https://www.sciencedirect.com/science/article/pii/S1094202513000471


https://www.federalreserve.gov/newsevents/speech/bernanke20100827a.htm


https://www.federalreserve.gov/newsevents/speech/bernanke20120831a.htm


https://www.federalreserve.gov/newsevents/pressreleases/monetary20081006a.htm


https://www.federalreserve.gov/newsevents/pressreleases/monetary20120125c.htm


https://www.federalreserve.gov/monetarypolicy/policy-normalization.htm

__________. 2017b. “Policy Tools: The Discount Rate.”
https://www.federalreserve.gov/monetarypolicy/discountrate.htm


Divisions of Research & Statistics and Monetary Affairs, September.


https://www.frbservices.org/assets/central-bank/tdf-overview.pdf


Craig, Ben, Sara Millingtion, and John Zito. 2014. “Who is Holding All the Excess...


https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr745.pdf


https://www.imf.org/~/media/Files/Publications/WP/2017/wp17129.ashx

https://www.brookings.edu/bpea-articles/the-zero-bound-on-interest-rates-and-optimal-monetary-policy/

http://www.nber.org/papers/w24039

http://www.ijcb.org/journal/ijcb15q1a8.pdf


https://www.newyorkfed.org/markets/opolicy/operating_policy_170524a

https://www.federalreserve.gov/monetarypolicy/o693lead.pdf

https://doi.org/10.17016/2380-7172.1875. 


Keister, Todd. 2012. “Corridors and Floors in Monetary Policy.” Liberty Street Economics,
Federal Reserve Bank of New York, April 4.


https://www.newyorkfed.org/research/epr/08v14n2/exesummary/exesum_keis.html


https://www.federalreserve.gov/newsevents/speech/kohn20091116a.htm


https://www.jstor.org/stable/2729790


Pozsar, Zoltan. 2016. “What Excess Reserves?” *Global Money Notes #5*, Credit Suisse, April 13. [https://research-doc.credit-suisse.com/docView?language=ENG&format=PDF&document_id=10600918n&sour ce_id=emcm&serialid=6kOK62k2brRsDf2wXpet090Fz7BQArYk93VobPupfmk=](https://research-doc.credit-suisse.com/docView?language=ENG&format=PDF&document_id=10600918n&sour ce_id=emcm&serialid=6kOK62k2brRsDf2wXpet090Fz7BQArYk93VobPupfmk=)


Reis, Ricardo. 2016. “Funding Quantitative Easing to Target Inflation.” In *Designing Monetary Policy Frameworks for the Future*. Kansas City: Federal Reserve Bank of
Kansas City, pp. 423-478.


__________. 2017b. “Interest on Reserves: A Secret Fiscal Weapon We’re Better Off

135
Without.” Alt-M, January 25. https://www.alt-m.org/2017/01/25/interest-on-reserves-a-secret-fiscal-weapon-were-better-off-without/


Walsh, Carl E. 1984. “Optimal Taxation by the Monetary Authority.” NBER Working


http://newmonetarism.blogspot.com/2017/02/balance-sheet-blues.html


https://www.federalreserve.gov/newsevents/testimony/yellen20170712a.htm