

DISCUSSION PAPER SERIES

DP15882

A MODEL OF QE, RESERVE DEMAND AND THE MONEY MULTIPLIER

Karl Whelan and Ellen Ryan

MONETARY ECONOMICS AND FLUCTUATIONS



A MODEL OF QE, RESERVE DEMAND AND THE MONEY MULTIPLIER

Karl Whelan and Ellen Ryan

Discussion Paper DP15882

Published 03 March 2021

Submitted 02 March 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Monetary Economics and Fluctuations

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Karl Whelan and Ellen Ryan

A MODEL OF QE, RESERVE DEMAND AND THE MONEY MULTIPLIER

Abstract

Quantitative easing programmes have driven unprecedented expansions in the supply of central bank reserves around the world over the past two decades, fundamentally changing the implementation of monetary policy. The collapse in money multipliers following QE episodes has often been interpreted as implying banks are happy to passively hold most of the reserves created by QE. This paper develops a simple micro-simulation model of the banking sector that adapts the traditional money multiplier model and allows for bank reserve demand to be inferred from monetary aggregates. The model allows the use of unwanted reserves by banks to play out over time alongside QE purchases and incorporates both significantly higher reserve demand after 2008 and capital constraints. With these additions, the model explains the persistently lower money multipliers seen in the US following QE, as well as the growth in commercial bank deposits. The model suggests the demand from banks for reserves has increased substantially since the introduction of QE but not to the point where banks are passively absorbing all newly created reserves.

JEL Classification: E51, E52, E58

Keywords: Quantitative easing, central banks, Money Multiplier

Karl Whelan - karl.whelan@ucd.ie
University College Dublin and CEPR

Ellen Ryan - Ellen.Ryan@ecb.europa.eu
European Central Bank

A Model of QE, Reserve Demand and the Money Multiplier

Ellen Ryan* Karl Whelan†

February 2021

Abstract

Quantitative easing programmes have driven unprecedented expansions in the supply of central bank reserves around the world over the past two decades, fundamentally changing the implementation of monetary policy. The collapse in money multipliers following QE episodes has often been interpreted as implying banks are happy to passively hold most of the reserves created by QE. This paper develops a simple micro-simulation model of the banking sector that adapts the traditional money multiplier model and allows for bank reserve demand to be inferred from monetary aggregates. The model allows the use of unwanted reserves by banks to play out over time alongside QE purchases and incorporates both significantly higher reserve demand after 2008 and capital constraints. With these additions, the model explains the persistently lower money multipliers seen in the US following QE, as well as the growth in commercial bank deposits. The model suggests the demand from banks for reserves has increased substantially since the introduction of QE but not to the point where banks are passively absorbing all newly created reserves.

*European Central Bank. ellen.ryan@ecb.europa.eu

†University College Dublin. karl.whelan@ucd.ie

The remarkable feature of the recent episode of quantitative/ credit easing was that this money multiplier totally failed to work [...]Why?

Charles [Goodhart \(2010\)](#)

1 Introduction

Prior to the global financial crisis, the framework underpinning monetary policy implementation was broadly agreed. The primary instrument of monetary policy was the short term interest rate. Central banks decided on a target rate and then set the supply of reserves to ensure that money market rates approximately matched this target rate. This changed with the onset of the crisis, as quantitative easing (QE) programmes, involving steady large-scale increases in the supply of reserves and thus the monetary base, became central to monetary policy toolkits around the world.

The traditional theoretical framework for understanding the effects of monetary expansion is the money multiplier model. In this model the broader supply of money is always a fixed multiple of the monetary base and this ratio is maintained by banks using up all excess reserves following any expansion of the monetary base. This model had fallen out of favour with practitioners of monetary policy long before the financial crisis, partly due to its highly simplistic assumptions about the behaviour of banks, ignoring the complexities caused by modern capital and liquidity regulations. In addition, decisions by some major central banks to begin paying interest on reserves were seen by some as invalidating the traditional model's assumption that banks would always seek to get rid of reserves in excess of legal reserve requirements. For example, an influential Federal Reserve study ([Keister et al. \(2008\)](#)) argued that the payment of interest on reserves would mean that banks would be indifferent between any quantity of reserves above the legally required amount.

The collapse in money multipliers around the world following the introduction of QE was taken by many as final proof that the traditional framework was defunct, with [Goodhart \(2010\)](#) arguing that *“it should be discarded immediately”*. Indeed, the literature on QE has placed little emphasis on the idea that banks seek to shift excess reserves off their balance sheets and expand the supply of credit by doing so. For example, in his summary of how QE affects the economy, [Bernanke \(2020\)](#) points to two mechanisms—a portfolio balance mechanism and a role in signalling forward guidance—but does not mention the idea of additional reserves creating new bank credit and expanding the broader money supply.¹

More recently, however, empirical work by [Ryan and Whelan \(2021\)](#), [Altavilla et al. \(2018\)](#), [Kandrac and Schlusche \(2021\)](#) and [Christensen and Krogstrup \(2016\)](#) all point to an active response by the banking system to the creation of reserves under QE, with banks seeking to reduce their reserve holdings via asset purchases or loans. This suggests that there may still be some room for the traditional textbook money multiplier mechanism in which an expansion of reserves affects the supply of credit and money.

¹The literature on QE has examined the policy's effect on a range of key economic variables such as long-term interest rates (see for example [Gagnon et al. \(2011\)](#), [Christensen and Rudebusch \(2012\)](#) and [Eser and Schwaab \(2016\)](#)), asset prices ([Joyce et al. \(2012\)](#)), macroeconomic aggregates such as GDP and inflation ([Baumeister and Benati \(2017\)](#) and [Gambetti and Musso \(2017\)](#)) and market participant expectations ([Christensen and Rudebusch \(2012\)](#) and [Ciccarelli et al. \(2017\)](#)).

This paper presents a micro-simulation model that simulates a banking system's response to monetary expansion and links the demand for reserves by banks to developments in monetary aggregates. The model is calibrated to match the creation of reserves by the Federal Reserve since 2008 and generates predictions for bank deposits and the money multiplier in the US that can be compared with actual outcomes. This allows us to infer unobservable bank behaviours from readily available monetary aggregate data.

Our model is based on the traditional money multiplier framework but incorporates a number of additional mechanisms to address some of its shortcomings. First, we make a different assumption about the timing of the banking system's circulation of reserves. The traditional textbook model sees banks that receive reserves loaning out a fraction of them, meaning reserves end up at other banks which then make subsequent loans. This "multiplication" process is assumed to happen fast enough to maintain relative constancy of the ratio of broad money to the monetary base. In contrast, in our model, a bank that receives reserves can offload them via making loans or bond purchases during that period but subsequent usage of these reserves by other banks has to wait until the next period. This means the multiplication process plays out more gradually over time. Second, we assume that the demand for reserves by banks can exceed formal reserve requirements and discuss evidence indicating that demand for reserves has increased significantly during the period since QE was implemented. Third, after establishing our core model, we also incorporate a simple form of capital constraints.

Allowing deposit creation to occur gradually over time implies there should be a sharp fall in the money multiplier following a large expansion in the monetary base, even when the banking system's demand for reserves remains unchanged. When our model is calibrated to US data, the initial money multiplier drop seen at the onset of QE can be largely replicated without assuming any change in bank behaviour. However, the persistently low money multipliers seen since 2008 indicate an increase in reserve demand in the banking system. Our analysis suggests that over the period 2008-2019, banks appeared to wish to keep reserves equivalent to 15-25 per cent of deposits. While this is a big increase in the demand for reserves relative to the pre-QE period, our analysis still implies there was a large amount of unwanted reserves during the post-QE period, suggesting banks were not *"happy just to sit on their massively increased cash base"* (Goodhart (2010)).

When we incorporate capital constraints, our model is able to broadly replicate dynamics in both the money multiplier and aggregate deposits. Moreover, the model begins signalling a shortage of reserves in 2019, which is roughly consistent with the reserve shortages that caused financial market disruptions in September of that year.

The rest of this paper is organised as follows. Section 2 describes the changing approach to money policy implementation in the US since the financial crisis, with a focus on the role of reserve balances. Section 3 lays out our model. Section 4 provides some simple illustrative outputs. Section 5 calibrates the model to the US using pre-QE monetary aggregate data and data on Federal Reserve purchases under QE. Section 6 incorporates capital constraints and Section 7 concludes.

2 Reserves and the Fed’s Monetary Policy

Prior to the 2008 global financial crisis, the Federal Reserve implemented monetary policy by varying the volume of reserves in the US banking system so that average rates in federal funds market aligned with its target interest rate. Official reserve requirements were low and the total stock of reserves in the system was also kept at a low level to ensure that money market rates remained sensitive to changes in the volume of reserves. This approach is illustrated in terms of reserve supply and demand in Figure 1a, taken from [Ihrig et al. \(2015\)](#). With the outbreak of the financial crisis the Federal Reserve embarked on a number of Large Scale Asset Purchase (LSAP) programmes, also referred to as QE. These programmes entailed a massive expansion in the outstanding stock of reserves. From about \$0.01 trillion in August 2008, total reserves reached \$2.8 trillion in 2014 (see Figure 2a).

In this setting, the traditional money multiplier model predicts an even larger expansion in broad money, as banks continuously used up excess reserve balances through lending and security purchases until broad money was equal to base money times the money multiplier. Instead, the M2 money multiplier, which had been relatively stable over the 15 years prior to the introduction of QE, halved in just a few months in late 2008 and early 2009 and fell further after other QE episodes before partially recovering after 2014. Figure 2b shows the measure of the money multiplier that is most relevant for our analysis. It adjusts M2 by removing the contribution of retail money market mutual funds. This allows us to focus solely on the contribution of the banking sector to developments in monetary aggregates.

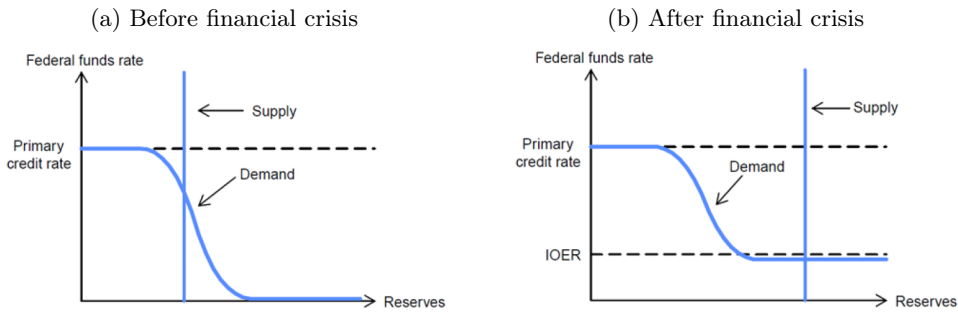
Much of the commentary around this unprecedented drop in the money multiplier focused on the idea that banks were happy to retain large reserve balances instead of reducing them through lending or security purchases. This was attributed to increased risk aversion following the financial crisis, stricter regulations applied by authorities and the decision by the Federal Reserve to begin paying interest on reserves ([Logan \(2019\)](#), [Ennis \(2014\)](#), [Keister and Andrews \(2009\)](#)).

After reaching their peak in 2014, total reserve balances began a gradual decline. In June 2017 the Federal Open Markets Committee (FOMC) announced plans to gradually reduce the size of the Federal Reserve’s balance sheet by ceasing reinvestment of principal payments from Treasuries and mortgage-backed securities (MBS), thus also reducing the outstanding stock of reserves.²

In this new ample reserve environment, the approach to policy implementation shown in Figure 1a was no longer feasible: The supply of reserves no longer intersected with the downward sloping part of the demand curve. This meant that the Fed moved to a framework in which the interest rate paid on excess reserve balances (IOER), a policy introduced in October 2008, became the key policy rate.

²The reduction began in practice in October 2017. For further detail see <https://www.federalreserve.gov/monetarypolicy/policy-normalization-discussions-communications-history.htm>.

Figure 1: Federal Reserve monetary policy implementation frameworks



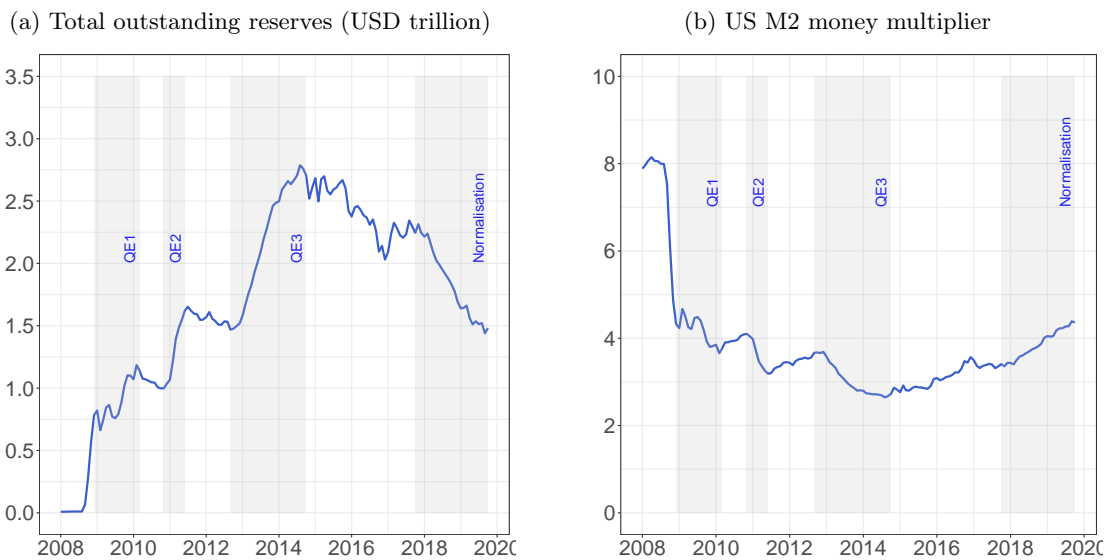
Note: The primary credit rate is the interest rate that the Federal Reserve charges banks to borrow overnight from its primary credit program as part of the Federal Reserve's discount window. Source:

[Ihrig et al. \(2015\)](#)

This approach to monetary policy is illustrated in Figure 1b, also taken from [Ihrig et al. \(2015\)](#). With a large supply of excess reserves, the interest rate paid on excess reserve balances (IOER) acts as a floor for the Federal Funds rate, as bank demand for reserves flattens out at this level.³ Thus when the supply of reserves is anywhere to the right of downward sloping section of the demand curve, the Federal Reserve can control short term rates by changing the IOER. The discount window's primary credit rate should act as an upper bound on the federal funds rate, creating a ceiling for the Federal Reserve's target range.

³To be precise, the Federal Funds rate flattens out just below the IOER because there are participants in money markets who are not commercial banks and don't get access to interest paid on reserves. See [Ihrig et al. \(2015\)](#).

Figure 2: Reserves expanded with the implementation of QE and the money multiplier dropped sharply



Note: M2 aggregates have been corrected for the contribution of retail money market funds - see Section 5.1 for discussion. Dates for QE programmes are taken from New York Fed website [here](#). The expansion in reserves in the month prior to QE1 is driven by Federal Reserve lending to banks following the outbreak of the financial crisis.

3 Model

We use a micro-simulation model to simulate the transmission of reserve expansion to monetary aggregates via the actions of banks. Banks maximise profits subject to a simple liquidity constraint which can be seen as banks' reserve demand arising from either regulatory constraints or risk aversion. By comparing our model output to observed dynamics in monetary aggregates, we can infer how this bank behaviour may have changed following the introduction of QE. Our micro-simulation approach also allows for us to model a realistic QE programme, incorporating heterogeneity in bank exposure to policy.

3.1 Introducing the model

The model consists of a banking system made up of n *ex ante* identical commercial banks and a central bank. In period $t = 0$, Bank i 's liabilities consist of deposits d_{i0} and its assets consist of reserves r_{i0} and other assets a_{i0} . In the first version of the model, we have $a_{i,t} + r_{i,t} = d_{i,t}$ so total assets always equal liabilities; equity capital funding is added later.

Banks are assumed to earn zero interest on reserves but earn a positive return on assets (ROA) on their holdings of other assets. We could add a "corridor" system in which banks earn interest on reserves but this would not change the dynamics of the model as long as reserves earn a lower return than other assets. Given these assumptions, banks seek to maximise profits subject to a liquidity constraint whereby other assets cannot exceed a share β of total deposits.⁴

This simple maximisation problem can be expressed as follows:

$$\begin{aligned} \text{Max } \pi &= (ROA) a \\ \text{subject to } r &\geq (1 - \beta)d \end{aligned} \tag{1}$$

In the traditional model $1 - \beta$ is assumed to be fixed and equal to the reserve requirement. Thus any reserves in excess of $(1 - \beta)d$ are "excess reserves". However, recent events suggest that banks do in fact want to hold reserves well above this specific regulatory requirement. Where reserve demand exceeds the reserve requirement, reserves in excess of $(1 - \beta)d$ could more accurately be called "*unwanted reserves*". We use this alternative term for the rest of our analysis.

We are purposely vague about the exact nature of the other assets that banks invest in. The traditional literature on the money multiplier emphasises that increased reserves lead to an expansion in the money supply because banks make additional loans and these loans create deposits. However, it is also the case that a bank using reserves to purchase, for example, a bond, will increase deposits: The person who sold or issued the bond will see a credit to their deposit account and this will expand the broad money supply in the exact same way as if they had been issued a loan.⁵ All that is required for our analysis is that,

⁴It is possible in the model for a bank to enter a period with this constraint not holding. We discuss below how banks are assumed to behave in this situation.

⁵As Goodhart (2017) points out "*A purchase by a commercial bank of public sector paper, a bond or a bill, creates money in exactly the same way as does a bank loan to the private sector.*" The writings of Friedman and Schwartz, the economists most associated with the concept of the money multiplier, also do

Table 1: Step 1 - central bank carries out QE purchases

(a) Step 1: Bank 1			
Reserves	$r_{10} + \frac{\Delta R_1}{mn}$	Deposits	$d_{10} + \frac{q\Delta R_1}{mn}$
other assets	$a_{10} - \frac{(1-q)\Delta R_1}{mn}$		

(b) Step 1: Bank 2			
Reserves	r_{20}	Deposits	d_{20}
other assets	a_{20}		

above a certain demanded level of reserves, banks would prefer to hold other assets.

3.2 Step 1: The central bank changes the size of the monetary base

The model runs over a finite number of periods and within each period a number of consecutive steps occur.

Asset purchases under QE result in an expansion of the monetary base, creating new reserves ΔR_t . In practice, banks can receive these reserves either by selling assets to the central bank themselves or as result of non-bank entities selling assets to the central bank. Empirical studies have found that non-bank entities and non-resident entities accounted for a high share of sales to central banks during QE episodes (ECB (2017), Carpenter et al. (2015)). Because these institutions do not have access to central bank deposit accounts, purchases are carried out via the domestic banking system which accepts reserves on behalf of non-bank sellers and then credits their deposit balances. This mechanism is reflected in our model where share q of purchases are carried out with non-banks. On aggregate the banking system gains ΔR_t in new reserves, $\Delta R_t q$ in new deposits and sells $\Delta R_t(1 - q)$ in assets itself to the central bank.

Heterogeneity across banks in their direct exposure to reserves created by QE has also been well documented during recent QE programmes (see Baldo et al. (2017) for the euro area). In our model, a share s of banks act as central bank counterparties, allowing for $1 - s$ to never receive reserves directly from the QE programme. In a given period, a share m of banks receive reserves and when $m < s$ these are randomly selected in each period. Each bank which does receive reserves gets an equal share: $\Delta R_t/mn$. When the central bank is reducing the stock of reserves, these dynamics occur in a symmetric fashion in the opposite direction.

Table 1 illustrates this step after the central bank’s first market intervention. Bank 1 is one of the mn banks to interact with the central bank and Bank 2 is one of the other $(1 - m)n$.

not assume deposit creation is caused only by loans. For example, Friedman and Schwartz (1987) discuss how an increase in reserves means that “it now has larger reserves than it has regarded before as sufficient and will seek to expand its investments and its loans at a greater rate than before.” They also noted (page 66) that banks were more likely to first use excess reserves to purchase securities and only later to turn to loans.

Table 2: Step 2 - Repayments on other assets are settled

(a) Step 2: Bank 1			
Reserves	$r_{10} + \frac{\Delta R_1}{mn} + w_{11} - z_{11}$	Deposits	$d_{10} + \frac{q\Delta R_1}{mn} - z_{11}$
other assets	$a_{10} - \frac{(1-q)\Delta R_1}{mn} - w_{11}$		

(b) Step 2: Bank 2			
Reserves	$r_{20} + w_{21} - z_{21}$	Deposits	$d_{20} - z_{21}$
other assets	$a_{20} - w_{21}$		

3.3 Step 2: Repayments on other assets are made

Banks' reserve, deposit and other asset balances are also affected by the actions of non-bank actors, such as borrowers engaging in loan repayment. For example, a borrower with deposits in Bank 1 and a loan from Bank 2 makes a loan repayment by asking Bank 1 to reduce their deposit balance and transfer this to Bank 2. Bank 1 will carry out the transfer using reserves. This reduces the deposits and reserves held by Bank 1 and increases reserves held by Bank 2. As the loan has been partially repaid, Bank 2's loan book shrinks by an equal amount. Coupon or principal payments on bonds owned by banks could be modelled in a similar fashion.

In our model, we assume a share θ of each bank's outstanding other assets are repaid in a given period. Deposit outflows caused by aggregate repayments are then distributed across banks in line with their share of aggregate deposits. This process is shown in Table 2, using $w_{i,t}$ to denote loan repayments and $z_{i,t}$ to denote deposit outflows, and also taking into account payments in the opposite direction (a borrower from Bank 2 repaying a loan using deposits from Bank 1).

3.4 Step 3: Banks optimise their asset portfolio

Because the return on reserves is lower than the return on other assets, a profit maximising bank will rebalance their portfolio so that reserves are at the lowest level required to meet their liquidity constraint. Commercial banks can remove reserves from their accounts by making loans or buying debt securities from the non-bank sector. The outflow of reserves ($o_{i,t}$) from each bank is calculated as:

$$reserve\ outflow = reserves - (1 - \beta)deposits \tag{2}$$

This is accompanied by a simultaneous and equal increase in holdings of "other assets". This is shown in Table 3 which assumes that $o_{1,2} = 0$ for the sake of simplicity.

Table 3: Step 3 - Banks optimise their asset portfolio

(a) Step 3: Bank 1			
Reserves	$r_{10} + \frac{\Delta R_1}{mn} + w_{11} - z_{11} - o_{11}$	Deposits	$d_{10} + \frac{q\Delta R_1}{mn} - z_{11}$
other assets	$a_{10} - \frac{(1-q)\Delta R_1}{mn} - w_{11} + o_{11}$		
(b) Step 3: Bank 2			
Reserves	$r_{20} + w_{21} - z_{21}$	Deposits	$d_{20} - z_{21}$
other assets	$a_{20} - w_{21}$		

3.5 Step 4: Reserve outflows for Bank 1 land in the account of Bank 2

These reserves do not just disappear: Reserves can only be held by banks with access to central bank accounts and therefore exist within a closed system. For example, Bank 1 may create a loan to reduce its reserve balance. This will result in the creation of a loan on the asset side of its balance sheet and a deposit in its liabilities. When the borrower uses the deposits to buy an asset, Bank 1 will settle this transaction by reducing the borrower's deposit balance and transferring an equal value of reserves to Bank 2 where the asset seller keeps their deposits. Bank 2 will then increase the deposit balance of the asset seller.

We model this by allocating each bank's reserve outflow to a random bank in the system, as shown in Figure 4 where Bank 2 receives reserve inflows from Bank 1. As this is the end of period, this balance sheet describes the situation banks will start the next period in. Crucially, this means that despite all banks rebalancing portfolios to achieve their optimal reserve balance, many will immediately find themselves holding unwanted reserves again. Note that we are assuming at this point that banks pay out all profits to shareholders as dividends but we will return to this issue later.

One point worth clarifying, is that it is possible for a bank to end the period (and thus start the next period) with lower than its desired amount of reserves, ($reserves < (1 - \beta)deposits$). We assume that the bank acts to restore its desired reserve holdings by not acquiring any further other assets and allowing loan repayments and deposit outflows to return them over time to their desired level of reserves.⁶

3.6 End of period: Calculate monetary aggregates

Once these steps have been completed the period ends and aggregate deposits (D), reserves (R) and other assets (A) can be calculated.⁷ The money multiplier is the ratio of the broad

⁶The model can be extended to allow for banks in reserve deficit to borrow reserves from banks with surpluses. However, this adds complexity without having a relevant impact on model results and so is not discussed here.

⁷As there are multiple steps within a given period t , each balance sheet item changes a number of times within a given period. We use the t subscript to denote end of period values.

Table 4: Step 4 - Bank 2 receives reserve outflows from Bank 1

(a) Step 4: Bank 1			
Reserves	$r_{10} + \frac{\Delta R_1}{mn} + w_{11} - z_{11} - o_{11}$	Deposits	$d_{10} + \frac{q\Delta R_1}{mn} - z_{11}$
other assets	$a_{10} - \frac{(1-q)\Delta R_1}{mn} - w_{11} + o_{11}$		

(b) Step 4: Bank 2			
Reserves	$r_{20} + w_{21} - z_{21} + o_{11}$	Deposits	$d_{20} - z_{21} + o_{11}$
other assets	$a_{20} - w_{21}$		

money supply (deposits plus currency) to the monetary base (reserves plus currency).

$$M_t = D_t + C_t \quad (3)$$

$$MB = R_t + C_t \quad (4)$$

$$Multiplier = \frac{M}{MB} = \frac{D_t + C_t}{R_t + C_t} \quad (5)$$

This means that to characterise the behaviour of the money multiplier, we need to make some assumption about how currency evolves over time in our model.

One approach would be to follow the traditional textbook approach and model currency as a “leakage” from deposits to finance non-bank-intermediated transactions, so that some fraction of each newly created deposit ends up as cash. We do not follow this approach for a few reasons. The first is that the demand for currency is largely unrelated to transactions demand. With currency in circulation in 2020 standing at over \$2 trillion, the average amount of currency in circulation per US resident is over \$6,000, which is clearly far larger than the amount being held for transactions purposes. Also, over three-quarters of the value of US dollars in circulation is accounted for by \$100 bills that have an average lifespan of 23 years.⁸ This suggests the use as a store of value and the facilitation of illegal activity account for most of the supply of US currency.⁹ The second reason is that we are modelling the effect of QE on monetary aggregates and it seems unlikely that investors who sold securities to the Fed then took out much of the proceeds in the form of cash.

For these reasons, the approach we take in our empirical calibration of the model is to model currency as evolving in line with the historical data for the currency-to-reserves ratio. In other words, we are not taking a stance on explaining this aspect of the monetary aggregates because we consider it to be outside the scope of the paper.

3.7 Comparison with traditional model

In the traditional money multiplier model, Steps 1, 3 and 4 repeat until the banking system produces enough deposits for the representative bank to have no unwanted reserves. This

⁸Evidence on the value of various denominations in circulation is available at [here](#) and evidence on average life spans is available [here](#)

⁹Indeed [Bernanke \(2010\)](#) noted that more than half of US currency by value was held outside of the US.

process is implicitly assumed to occur within a single period, or at least to complete itself before the central bank carries out another open market operation. As a result, we have a proportional relationship with

$$M_t = \frac{MB_t}{1 - \beta} \quad (6)$$

where the constant $1/1 - \beta$ is the money multiplier and $1 - \beta$ is typically assumed to be the reserve requirement.

Our approach differs in a number of ways. First, we allow for this process to play out over time and for QE purchases to happen alongside the banking system's absorption of new reserves through deposit creation. This small and intuitive change has important implications for the response of monetary aggregates to QE, as we will illustrate in the next section.

Second, is our characterisation as currency in circulation as evolving independently of bank deposits, as just discussed. The textbook approach to the money multiplier usually assumes there is a constant ratio of currency in circulation to deposits.

Third, the addition of loan repayment dynamics also means that individual banks do not stop creating new loans and deposits once they reach their desired level of reserves. Instead, the system as a whole reaches equilibrium when aggregate unwanted reserves are equal to aggregate loan repayments. We define equilibrium as the point where monetary aggregates remain constant over time. Annex A shows that equilibrium is now reached when deposits are:

$$D_t = \frac{(1 + \theta\beta)R_t}{1 - \beta + \theta\beta} \quad (7)$$

Finally, we produce monetary aggregate dynamics by aggregating over individual banks and allow for QE to be conducted via purchases from non-bank entities. This allows us to create a more realistic picture of a modern QE programme, including replicating the the market structure and bank-level frictions that exist around OMOs.

4 Illustrative Model Output

To demonstrate the dynamics our model produces, we start with some simple stylised examples. We first examine the response to a single period of QE purchases. Figure 3a provides the response of the money multiplier when the model is calibrated as follows:

- **Banking system structure:** The banking system is made up of 100 *ex ante* identical banks ($n = 100$). All banks are able to interact with the central bank ($s = 1$) but only half in given a period ($m = .5$).
- **Bank balance sheets and behaviour:** All banks want to hold reserves equal to one per cent of deposits ($\beta = 0.99$). In $t = 0$ total deposits held by the banking system equal 100. To ensure the system begins in equilibrium, as described by Equation (7), initial aggregate reserves are backed out of Equation 7 and come to 2.92 (rounded to two decimal places). As assets equal liabilities in all periods, initial aggregate other assets come to 97.08. 2 per cent of other assets are repaid each period.¹⁰
- **QE programme:** The central bank carries out no purchases for $t = 1$ to $t = 5$. In period $t = 6$ the central bank carries out a once-off purchase which increases the supply of reserves by 1.67 and then ceases purchasing activity again. In all periods three quarters of entities selling assets to the central bank are non-banks ($q = 0.75$).

In our first version of this simulation, we model the aggregate currency to reserve ratio as being constant at 5:1. Technically, this means the monetary base increases by 10 units in period 6, with one sixth of this increase being reserves and the rest being currency. While this is not a realistic representation of how the monetary base expands during QE, we use it as an initial baseline because it generates a constant long-run money multiplier. Specifically, this calibration implies a constant ratio of currency to deposits of 14.61 and an equilibrium money multiplier that can be calculated as 6.54 as follows:

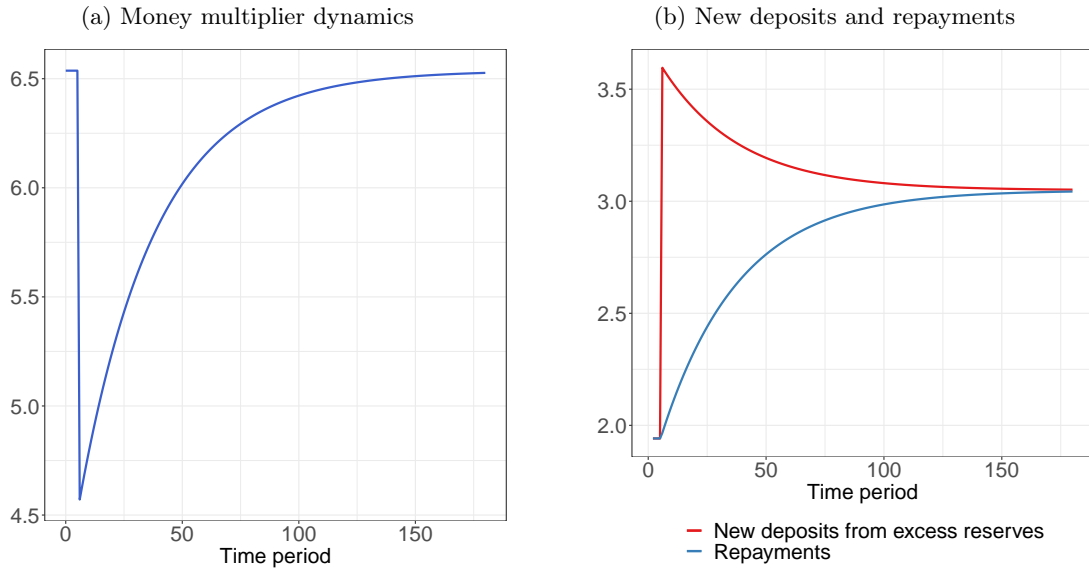
$$money\ multiplier_0 = \frac{D_0 + C_0}{R_0 + C_0} = \frac{100 + 14.61}{2.92 + 14.61} = 6.54 \quad (8)$$

In each period, deposits are reduced by total loan repayments but then increased again as banks use up unwanted reserve balances and create new deposits. Loan repayments and excess reserves are equal in size in period 1, so the system begins in equilibrium and remains in equilibrium until central bank purchases begin.

In period 6 the central bank expands the monetary base by 10 units and we see an abrupt drop in the money multiplier from 6.54 to 4.47. The drop has occurred without any change in bank behaviour and with each individual bank continuing to use all unwanted reserves to create new deposits, just as they would in a traditional money multiplier model. The main driver of this drop is the banking system's inability to create enough deposits to use up the *system's* unwanted reserves in a single period. In Period 6 the banks receiving new reserves from the central bank jointly create 2.61 units of new deposits in using up their individual unwanted reserve balances. Other banks create 0.99 units as they begin the period with some unwanted reserves. However, 57 units of new deposits are needed

¹⁰All numbers in this section are rounded to two decimal points.

Figure 3: Illustrative model output - one period of QE



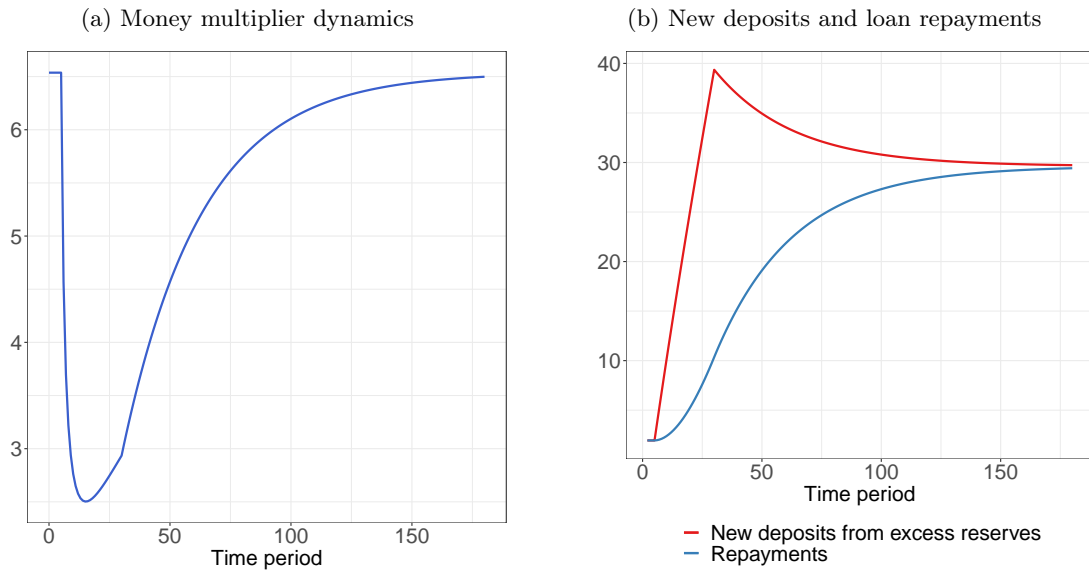
to return the system to equilibrium after the increase in reserves. These are created over time through an iterative process and the multiplier gradually recovers.

Figure 3a shows that this recovery can in fact be very slow. In each subsequent period repayments are made and banks then use up any unwanted reserves by increasing their holdings of other assets (creating loans or buying new debt securities), creating new deposits in the process. Figure 3b shows the system slowly returning to equilibrium as deposit growth reduces unwanted reserves and growth in the stock of other assets increases the size of repayments each period. The system, and as a result the money multiplier, returns to equilibrium when new deposits created each period equal repayments. For this calibration the process takes over 100 periods. Return to equilibrium is faster when β is lower but is still far from immediate.

To understand the dynamics produced by a QE programme in which consistent purchases are made over a period time, we run the model with continuous purchases of 5 between periods 6 and 30. Results in Figure 4a show a similar sharp drop following the introduction of the programme. As purchases continue the money multiplier continues to fall and reaches a low of 2.50 by period 15. From period 15 until the end of the QE programme, the money multiplier gradually rises but remains substantially below its pre-QE level. When purchases end, this process speeds up but the multiplier does not return to approximately pre-QE levels until after period 150.

This suggests that the sharp drop in money multipliers around the world following the implementation of QE does not necessarily invalidate the money multiplier model. It may simply be that the banking system needs time to fully react to a large increase in reserves. Where reserves are continuously introduced as this absorption is still taking place, a persistently lower money multiplier may also be in line with the model's fundamental mechanics. However, the size and persistence of these drops may also reflect changing behaviour by the banking system. Specifically, β may have fallen as banks began holding reserves equivalent to a higher share of deposits. In the next Section we aim to disentangle

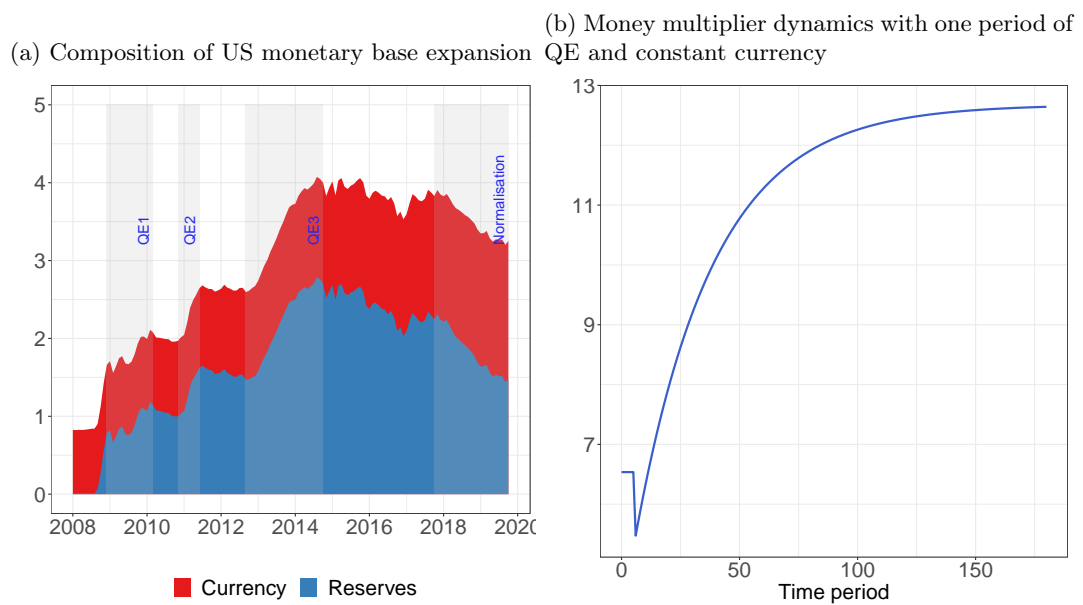
Figure 4: Illustrative model output - continuous QE



these two effects we need to calibrate our model to match real world data.

Of course the assumption of a constant currency to reserve ratio is unrealistic. Figure 5a shows that US currency in circulation has been largely unaffected by the huge expansion in the monetary base accompanying QE. Figure 5b shows that when currency is held fixed and the entire monetary expansion occurs via reserves, the model returns to equilibrium with a much higher money multiplier. This occurs because currency exists outside the banking system while reserves can be “multiplied”, so an increase in the ratio of reserves to currency leads to a higher equilibrium money multiplier. This means changes in the supply of currency are another element that can affect the dynamics of the money multiplier. We will control for these movements in our empirical calibration to distinguish between movements in the money multiplier that come from this source rather than changes in the supply of reserves or changes in bank behaviour.

Figure 5: Implications of changing currency to reserve ratios



5 Matching the US Money Supply

We calibrate our model to the US banking system by matching starting values for deposits and the monetary base to those observed prior to QE. Our model central bank then carries out purchases in line with the Federal Reserve’s LSAPs and we examine the model’s predictions for the broad money and the money multiplier for a range of values for β .

5.1 Calibration of parameters

We set starting aggregate bank deposits equal to the August 1st 2008 value of M2 subtracting currency in circulation and retail money market funds. This deposit aggregate is also used when calculating the money multiplier. The initial monetary base is calculated as the sum of currency in circulation and the total value of reserves held by banks with the Federal Reserve on the same date. Federal Reserve purchases are then captured by month-on-month changes in the size of the monetary base.

We run our model with a monthly frequency which should roughly reflect the average amount of time taken to use up reserves through new loan creation or through security purchases. The Federal Reserve’s open market operations are performed exclusively with a specific set of institutions also known as “primary dealers”. However these typically act as conduits to the rest of the financial system and related settlements should easily occur within the monthly time frame that we run our analysis on. As such, the bank receiving reserves from the central bank in our model is equivalent to the bank holding reserves for the ultimate counterparty as opposed to the primary dealer. Shorter frequencies may not be sufficient to allow for settlement with the ultimate seller and subsequent reserve outflows. That said, in practice, re-running the model with a weekly frequency generates very similar results to those we report here.

To ensure we maintain the correct composition for the monetary base throughout, we allow the currency to reserve ratio to vary over time in line with observed values. Our model is run from August 2008 to October 2019: We have omitted data from 2020 because the circumstances surrounding the provision of liquidity by banks during the pandemic are likely to represent unique circumstances that would be difficult to model. In the later periods of our sample, changes in the monetary base also capture the Federal Reserve’s efforts to shrink their balance sheet and reduce total outstanding reserves. We use monthly data taken from the St. Louis Fed’s FRED database.

To calibrate the share of Federal Reserve purchases from entities outside the banking system (q), we make reference to [Carpenter et al. \(2015\)](#). Their analysis uses Flow of Funds data to examine the relationship between changes in the holdings of Treasuries and MBS by a range of sectors and changes in Federal Reserve balance sheet composition. They find that ultimate sellers were typically hedge funds, insurers, broker-dealers and investment funds.¹¹ However, we cannot rule out *any* selling activity by the banking system and so set the share of non-bank sellers at 90 per cent. In [Appendix C](#) we examine the effect of changes to this parameter and find it has little effect on the model’s predictions about the evolution of the money multiplier.

¹¹Technically they find that “households” are the most important selling sector but conclude that this is likely due to hedge funds being included in this category.

Given the broad range of selling sectors we allow for 80 per cent of banks to receive reserves from the Federal Reserve at some point (s). The remaining 20 per cent can be thought of as banks with purely retail deposit bases. To account for some degree of random monthly variation in this flow at the bank level, 60 per cent of banks (or three quarters of the total selling sample) receive reserves in a given period (m).

We approximate the repayment rate of other assets on the basis of the composition of US bank credit, including both loans and securities.¹² By assigning approximate average maturity values for each part of banks' portfolios we get a weighted average maturity of 7 years. This gives a monthly repayment rate of 1 per cent ($\theta = 0.01$). As in Section 4, our banking system comprises 100 identical banks. In Appendix C we show that this parameter has no effect on the evolution of the money multiplier.

5.2 Monetary aggregates and inferring bank behaviour

First we run our model assuming that banks do not change their behaviour following the introduction of QE. Specifically, we set β equal to 0.99 to reflect a 1 per cent reserve requirement, consistent with the very small level of reserves prior to the crisis. As in our illustrative model, the money multiplier drops sharply following the introduction of the simulated QE. In fact, Figure 6 shows that with this calibration we can explain about two thirds of the initial drop in the US money multiplier during late 2008.¹³ However, this calibration cannot account for the full initial drop and the simulated money multiplier recovers much more quickly than in reality. This suggests that the banking system did change the way it treated its reserve balances in the period following the introduction of QE.

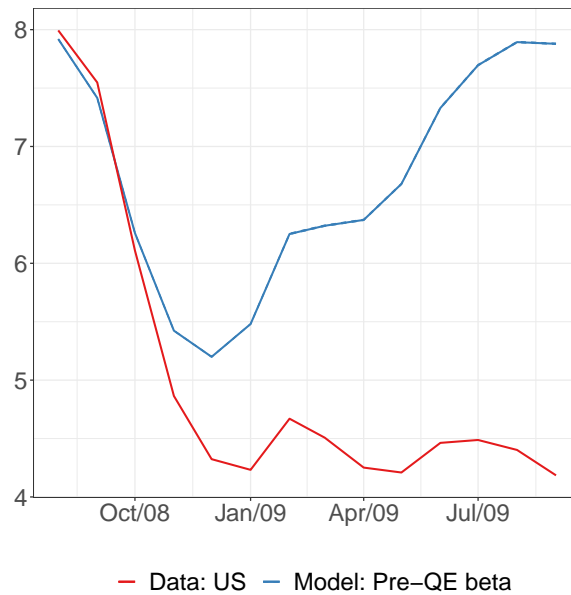
There are a number of reasons why banks have increased their demand for reserve balances with the Fed. During the global financial crisis, the demand for liquid assets increased due to precautionary demand. In later years, the regulatory environment that followed the Basel 3 agreement also increased demand. In particular, Basel 3 saw the introduction of a liquidity coverage ratio requiring banks to hold a sufficient amount of high-quality liquid assets to allow them to survive a period of significant liquidity stress lasting 30 calendar days. [Cecchetti and Schoenholtz \(2019\)](#) argue that the Fed's implementation of the liquidity coverage ratio regulations have particularly increased the demand for reserves because they are seen as superior to Treasuries when coping with a bank run, most notably because it allows withdrawals to be dealt with without the bank undertaking significant visible assets sales.

The Federal Reserve was, of course, aware of the increased demand for reserve balances but appears to have been surprised by the extent of the increase. Over a number of years in its annual report on open market operation, the New York Fed published estimates of the "normal" amount of reserves that would be consistent with meeting demand for reserves. These estimates rose from \$100 billion in 2015 to \$500 billion in 2017 and \$600 billion

¹²See Assets and Liabilities of Commercial Banks in the United States - H.8 [available here](#).

¹³The random allocation of reserves across banks following central bank actions and reserve outflows creates some very small random variation in model outcomes. To ensure that this is not driving results the model is run 100 times for all results shown in this section. Median, minimum and maximum outcomes for each period are plotted in all charts. However, as variation across iterations is limited the differences between these outcomes is not visible.

Figure 6: Model output with pre-QE bank behaviour compared to US money multiplier



in 2018. In March 2019, St Louis Fed economist Jane Ihrig (2019) noted that “*market participants have been mentioning estimates of the level of ample reserves to be in a range of \$1 trillion to \$1.4 trillion*”.¹⁴

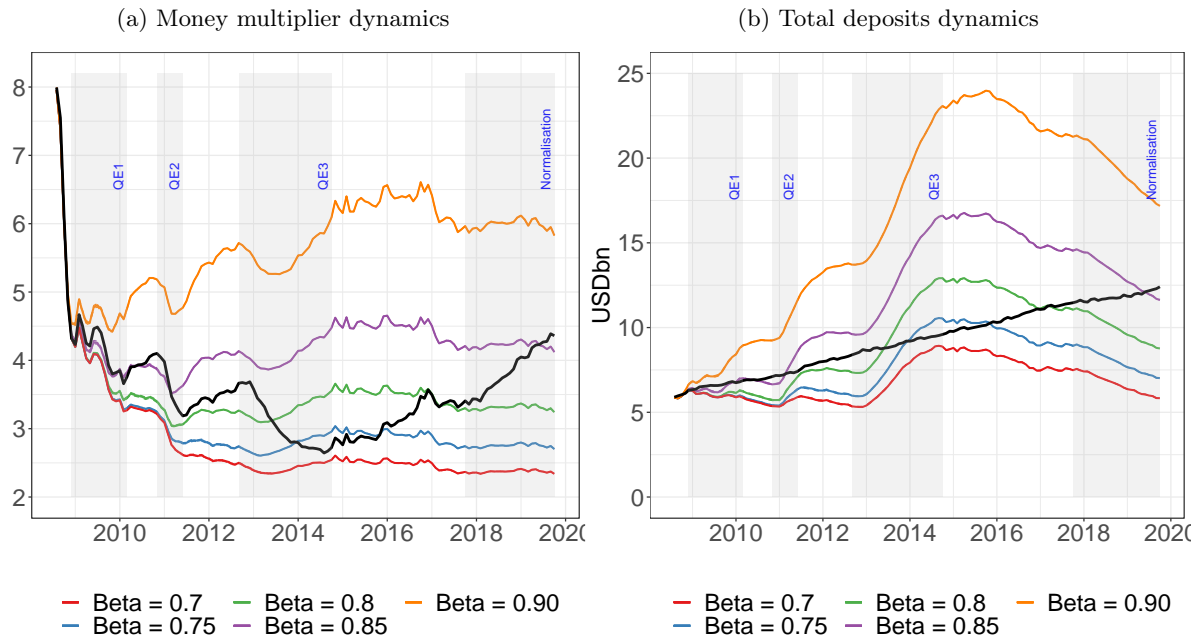
These much larger estimates were published in the context of a decision in January 2019 by the Federal Open Market Committee (FOMC) to continue operating an “ample reserves” policy, so that the supply of reserves would be well above demand. However, the Fed continued to contract the supply of reserves via its balance sheet normalisation policies. This ended in September 2019, with reserve balances reaching about \$1.4 trillion, following severe disruption in US repo markets which was ultimately linked to shortage of reserves relative to aggregate bank demand (Copeland et al. (2020)). Even with reserves well in excess of official reserve requirements, the banking system began to act as if it was in reserve deficit. In the context of our model, this level of reserve demand is consistent with a value of $\beta = 0.85$, meaning banks wish to retain 15 percent of their deposit base in the form of reserves.

In light of these events, we re-run the model with a range of different (but constant) β values. Figure 7a shows that the model does a reasonable job of replicating observed money multiplier dynamics (black line) with β values between 0.85 and 0.75. The model with $\beta = 0.75$ does a reasonable job of tracking the money multiplier’s behaviour up to 2014, while the final data point in our sample in 2019 is closer to the final value from the simulation with $\beta = 0.85$. In terms of bank behaviour, these results can be interpreted to mean that banks that previously kept almost no reserves on their balance sheets began keeping reserves equivalent to 15-25 per cent of deposits. While this is a huge increase relative to the pre-QE period, with these values of β , they still imply that banks had large unwanted reserve balances which they were actively pushing off their balance sheets.

The model is not as effective when it comes to replicating developments in total deposits at commercial banks. As before, Figure 7b shows that developments in US commercial bank

¹⁴Ihrig (2019)

Figure 7: Model output with a range of β values (coloured lines) compared to US data (black line)



Note: Black line shows dynamics from US data

deposits largely lie within the paths generated by the model output for $\beta = 0.85$ and $\beta = 0.75$. However, model output is much more volatile than the data, in particular it is much more responsive to changes in total reserves. This is to be expected given the simplicity of the model. In reality, banks may smooth loan creation over time, instead of responding immediately to changes in their reserve balances as they do in our model. Smoothing in loan creation could be due to credit demand, regulatory limitations and banks' desire to maintain good borrower relationships. Because loans create deposits, smoother loan provision would smooth out the development of aggregate deposits. We illustrate this by adding capital constraints to our banks.

6 Adding Capital Constraints

6.1 Adding capital to the model

We add capital (k) to the liability side of bank balance sheets and apply a capital constraint (x) which determines the maximum ratio of other assets to capital. The capital ratio is implemented in the model as a simple leverage ratio with the exception of allowing a zero risk-weight for balances kept at the central bank. Effectively, it is a constraint on the ratio of capital to other assets. An additional step is added to the end of each period, in which banks choose to distribute profits to shareholders or to use retained earnings to expand their capital base. When banks expand capital this new funding is immediately used to acquire new other assets so that at all times $a_{i,t} + r_{i,t} = d_{i,t} + k_{i,t}$.

Banks now maximise return on equity, subject to liquidity and capital constraints. This means they will only use retained earnings to expand capital when they are capital constrained. Banks solve the following problem:

$$\begin{aligned} \text{Max } ROE &= \frac{(ROA)a}{k} \\ \text{subject to } r &\geq (1 - \beta)d \\ \text{and } k &\geq \gamma a \end{aligned} \tag{9}$$

where γ is the minimum risk-weighted capital ratio for the bank. This could be interpreted as a regulatory requirement. However, it could also represent the desire by banks to operate with capital levels that are above minimum regulatory levels. Obviously, this treatment of the measurement of risk-weighted assets here—two assets, one with a risk weight of zero and the other with a risk weight of one—is a dramatic simplification of reality. Nevertheless, it turns out that even this simple approach dramatically improves the model’s ability to fit the data.

Because reserves still have a return lower than other assets, banks still aim to hold as few reserves as possible. However, the addition of a capital constraint puts an upper limit on the amount of other assets a bank can acquire in a given period. Specifically, banks can only expand their holdings of other assets up until the point where they equal $\frac{k}{\gamma}$. Reserve outflows can now be expressed as

$$o = \begin{cases} r - (1 - \beta)d, & \text{if } r - (1 - \beta)d + a < \frac{k}{\gamma} \\ \frac{k}{\gamma} - a, & \text{otherwise} \end{cases} \tag{10}$$

When banks are capital constrained, reserves previously classified as “unwanted” will remain on their balance sheet. As retained earnings are used to expand the capital base over time, these unwanted reserves will eventually be substituted for other assets.

Appendix B shows that the model now reaches equilibrium when:

$$D_t = \frac{(1 + \theta\beta)R_t - \theta\beta K_t}{1 - \beta + \theta\beta} \tag{11}$$

where K_t represents aggregate capital.

6.2 Illustrative output

To examine the effects of capital constraints on model dynamics we re-run our one-period QE model from Section 4. Figure 8 shows results for two simulations, one with a capital constraint of 8 per cent and one without a capital constraint. In both simulations banks begin with deposits of 100 and capital of 10. To ensure the system begins in equilibrium we derive starting reserves from Equation 11.¹⁵ Other assets have an ROA of 2 per cent and reserves pay interest of zero.¹⁶

The random allocation of reserves across banks following central bank interventions and reserve outflows from banks creates some random variation across outcomes. To ensure that results are not driven by random outcomes the full model is run 100 times for each calibration shown in this Section. Maximum, minimum and median outcomes are calculated for each variable in each period. The gap between minimum and maximum outcomes is shown by a shaded area and the median outcome is shown with the plotline. In most cases cross-iteration variation is negligible and the gap between minimum and maximum outcomes is not visible.¹⁷

Figure 8 shows that dynamics in both simulations are identical until banks with capital constraints (red line) hit their constraint. Constrained banks cannot substitute reserves for other assets as much as they would like and new deposit creation falls, slowing money multiplier growth. The constrained banking system returns to equilibrium with a marginally lower multiplier than the unconstrained system. This reflects the negative K entry in Equation 11 and the need for constrained banks to expand K to keep producing deposits.

Implications for more realistic QE scenarios become clear if we run repeated, large QE programmes through the new model. Figure 9 shows model output with three periods of QE, each consisting of 20 periods of 50 unit purchases and followed by a break of 50 periods.¹⁸ Here the unconstrained banking system produces aggregate deposit dynamics which are highly responsive to central bank purchasing activity, like our US calibrations in Figure 7b. The addition of capital constraints smooths these dynamics to an almost straight line, similar to that seen in US data. This is because capital constraints force banks to smooth deposit creation over time, instead of creating deposits in bursts aligned with the timing of the QE programme.

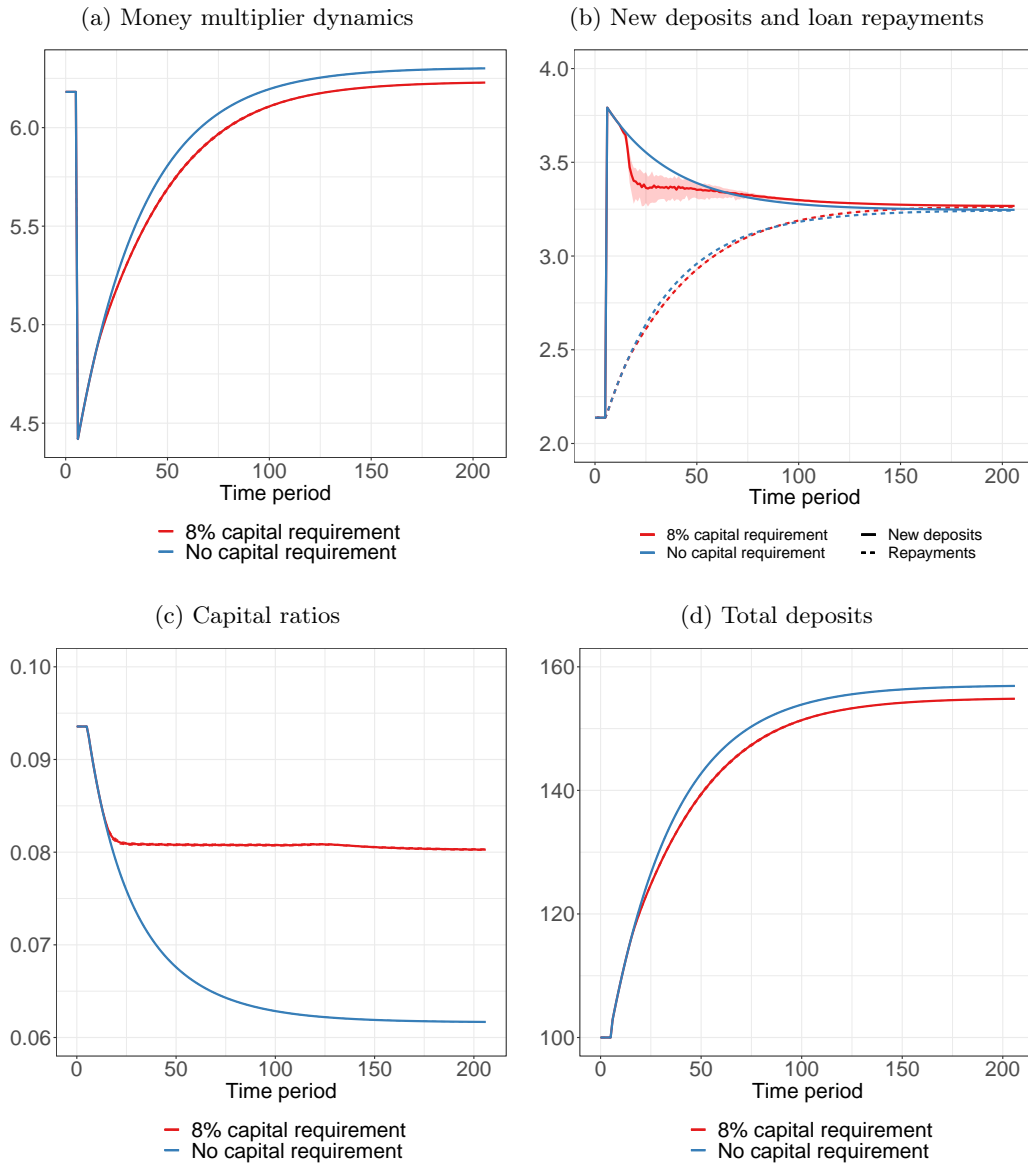
¹⁵Giving starting reserves of 3.12, and starting other assets of 106.88. This means all banks begin with risk weighted capital ratios of 10.67 per cent. This is for illustrative purposes - allowing banks to begin with excess capital allows us to see the point where they hit their capital constraint. If banks are maximising ROE they should never have excess capital.

¹⁶We would obtain the same outcome by assuming a different value for the interest rate on reserves, provided the net interest margin is kept at 2 percent.

¹⁷This was also the case in Section 4. However, cross-iteration variation was negligible and the difference between minimum, maximum and median outcomes is not visible in charts.

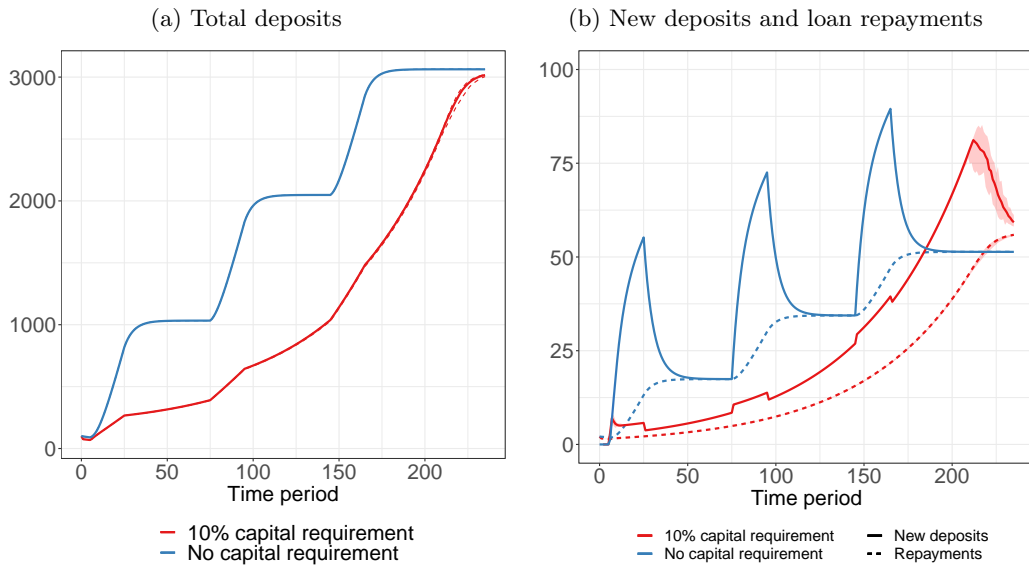
¹⁸ β is also reduced to 0.85 and starting capital reduced to 8.

Figure 8: One period of QE with capital constraints



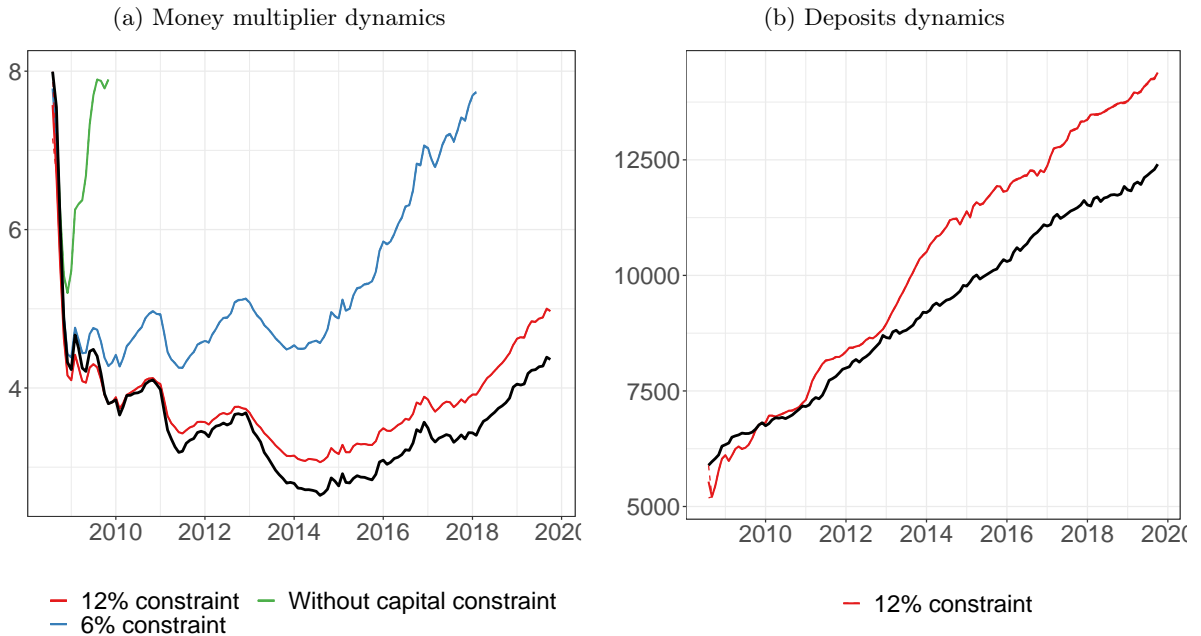
Note: The model is run 100 times for both calibrations. The gap between minimum and maximum outcomes is shown by a shaded area and the median outcome is shown with the plotline. In most cases cross-iteration variation is negligible.

Figure 9: Dynamics with repeated QE



Note: The model is run 100 times for both calibrations. The gap between minimum and maximum outcomes is shown by a shaded area and the median outcome is shown with the plotline. In most cases cross-iteration variation is negligible.

Figure 10: Adding simple capital constraints to US calibration with pre-QE β (0.99)



Note: Black line shows dynamics from US data. On LHS model output is only shown up until the point where the multiplier returns to its pre-QE levels. This is simply to enhance chart readability as output with no capital constraint continues to rise after 2010.

6.3 Calibration to US Data

Here, we incorporate the simple capital constraint into our US calibration from Section 5.1. As before the system begins with total deposits and total reserves equal to those held by US banks in August 2008. The model is run with capital constraints of 6 and 12 per cent, in line with aggregate core equity tier one ratios (expressed relative to risk-weighted assets) of US bank holdings companies at the start and end of the period examined. In all cases banks begin the simulation holding exactly the required amount of capital. ROA is set at 1 per cent, in line with average ROA for US bank holding companies over the period examined.¹⁹

Figure 10 shows model output with a β of 0.99 (our pre-QE beta from Figure 6). Model output is only shown up until the point that the multiplier returns to pre-QE levels. This is simply to enhance the readability of the chart as the multiplier with no capital constraint continues to rise from 2010 onward. With either capital constraint the model explains the entire initial drop in the multiplier. However, with the 6 per cent capital constraint the money multiplier returns to pre-QE levels by 2018. Introducing the stricter 12 per cent requirement allows the model to track US dynamics quite well, given the high β value. However, model output is still substantially higher than the US multiplier and provides a poor match to deposit dynamics, over-producing deposits from 2013 onward.

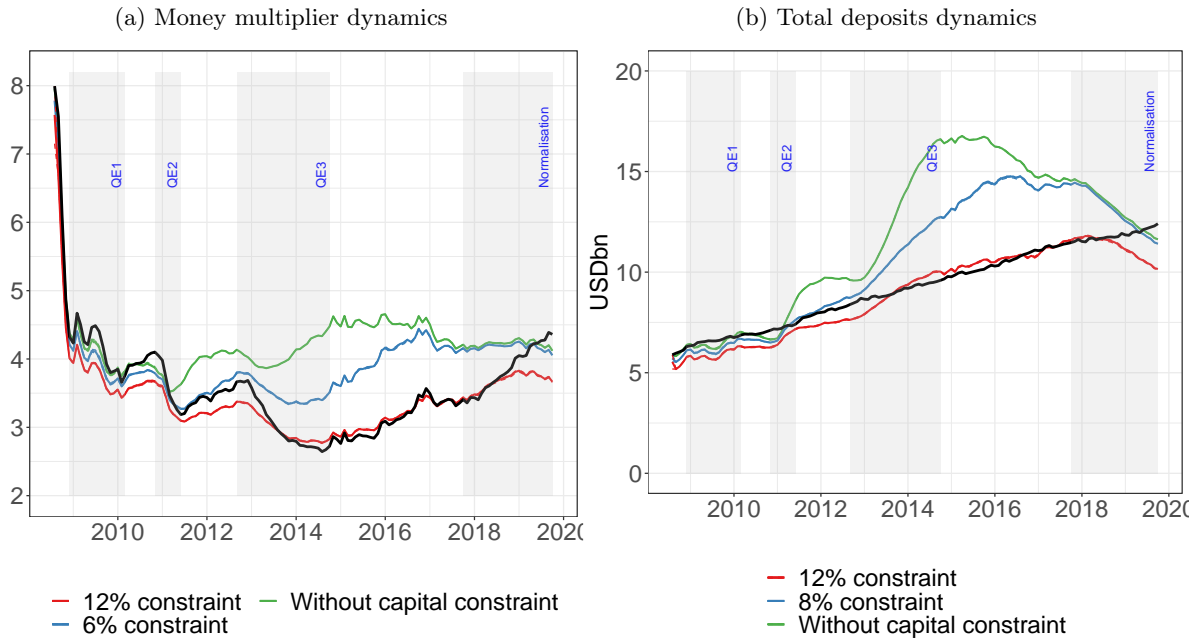
Next we re-run our model with a β of 0.85. Figure 12 shows that the 6 per cent capital requirement provides a good fit from 2008-2013. The 12 per cent capital constraint provides an almost perfect fit of both multiplier and deposit dynamics between 2014 and early 2019. The 12 percent constraint version of the model fitting better makes intuitive sense because the rising capital ratios of this period are consistent with banks targeting higher capital ratios than prevailed in 2008.

Finally, the 12 per cent constraint calibration suggests the US banking system was running low on reserves from early 2019 onwards. In our model, this is signalled by drops in both the multiplier and total deposits in the simulated model. While this deposit contraction did not occur, the model's parameterisation to reflect banks seeking to hold about 15 percent of their deposits in the form of reserves does appear to be backed up by the events of September 2019.

One question is whether this improved fit of deposit dynamics stems purely from the addition of a capital constraint and is not sensitive to the value of β . To illustrate that this is not the case, Figure 12 also includes results from simulations in which there is a 12 percent capital constraint and $\beta = 0.2$, i.e. a situation in which banks are happy to retain high quantities of reserves. The figure shows that this version of the model performs poorly in explaining deposit dynamics and the money multiplier. Put simply, the growth in deposits over the period since 2008 has not been consistent with banks passively absorbing reserves.

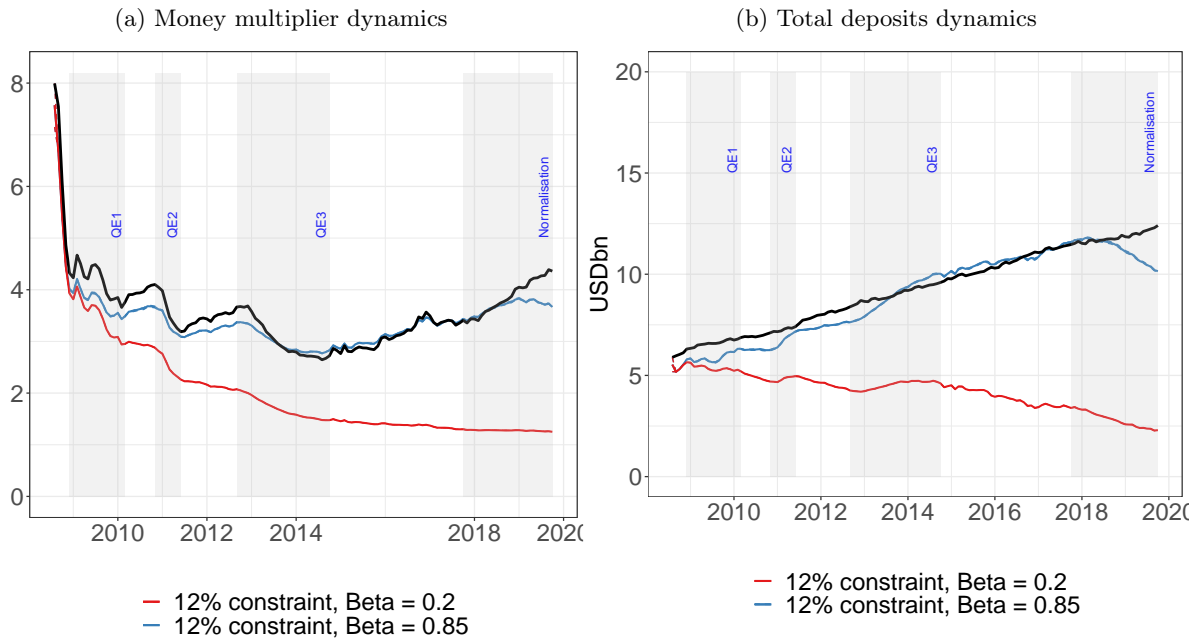
¹⁹See [Quarterly Trends for Consolidated U.S. Banking Organizations](#)

Figure 11: Adding simple capital constraints to US calibration with β of 0.85



Note: Black line shows dynamics from US data.

Figure 12: US calibration with capital constraints and β of 0.2



Note: Black line shows dynamics from US data.

This brings us back to the failure of the traditional money framework model following the introduction of QE. In 2010, Charles Goodhart suggested the model was failing to explain the behaviour of the money supply because banks were just happy to sit on their stock of reserves. Assuming that a significant increase in the demand for reserves likely took place even prior to the Basel 3 agreement concluded in 2010, he may well have been correct about how banks viewed the initial large increases in the supply of reserves. However, our calculations suggest there has been far too much deposit creation in the years since 2010 to be consistent with banks continuing to passively absorb reserves. While the textbook version of the money multiplier did fail, a version of the model that incorporates some simple and intuitive additions (allowing deposit creation by the banking system to play out over time as well as increased reserve demand by banks and capital constraints) can explain the observed dynamics quite well up until the period of reserve shortages during 2019.

7 Conclusions

Robert Solow (1956) famously said “*All theory depends on assumptions which are not quite true. That is what makes it theory.*” However, his work showed that macroeconomic models based on highly simplifying assumptions can still be useful if they provide a convenient framework to explain or summarise the relevant data.

Despite still featuring in most macroeconomic textbooks, the collapse in money multipliers following QE programmes over the past decade suggested the highly simplified money multiplier model is no longer a useful framework. This has raised the question of whether the behavioural assumptions underlying this model are completely false. In particular, if banks are happy to passively accept any amount of reserves supplied by central banks, then the mechanisms emphasised in the model will simply be absent in the real world. Indeed, the literature on the impact of QE has largely ignored the model’s predictions that an increase in reserves will trigger a corresponding expansion in bank credit and the broader money supply.

Our analysis suggests, however, that with a number of realistic alterations, a version of the money multiplier model can perform relatively well at explaining the money multiplier and total commercial bank deposits in the US since 2008. Factoring in the time it may take banks to create deposits in response to a change in reserves, modelling changes in the demand for reserves and adding a role for regulatory risk-weighted capital requirements, taken together, produce simulated aggregate series that matched the data for the last decade relatively well.

For central bankers, there are a number of potential policy implications. A first is that the short-run failure of the banking system to create credit in response to large increases in reserves doesn’t necessarily imply there is no longer-run link between the supply of reserves and bank credit. Our simulations suggest a large increase in the supply of reserves could produce a long and gradual process of credit creation. It also suggests policies to reduce the supply of reserves can eventually have a negative impact on bank credit. A second is that central banks should not assume that the payment of interest on reserves automatically means reserves are passively absorbed by the banking system. A third is the importance of having accurate estimates of the underlying demand for reserves in the banking system. While there have been regular exercises over the past decade to do this, the events of September 2019 illustrate that this demand has been sufficiently unpredictable to catch the Federal Reserve by surprise.

Our model opens up a range of possibilities for further work on the behaviour of banks in an environment of substantial excess reserves. Aggregate dynamics discussed in this paper are calculated from a bank-level panel produced by our model. This allows for the introduction of different bank types that could vary in terms of liquidity and capital constraint. Matching the distribution of reserves across banks to those seen in a micro data set could also provide insight into reserve shortages in particular parts of the banking system, another potential cause of the September 2019 disruption. While we examine the effects of liquidity and capital constraints, further work could incorporate their drivers. This could include further examination of the demand for reserves across different types of banks and in different market environments.

References

- Altavilla, Carlo, Miguel Boucinha, Sarah Holton, and Steven Ongena**, “Credit Supply and Demand in Unconventional Times,” Working Paper 2202, European Central Bank, 2018.
- Baldo, Luca, Benoit Hallinger, Caspar Helmus, Niko Herrala, Débora Martins, Felix Mohing, Filippos Petroulakis, Marc Resinek, Olivier Vergote, Benoit Usciati, and Yizhou Wang**, “The Distribution of Excess Liquidity in the Euro Area,” Occasional Paper Series 200, European Central Bank, November 2017.
- Baumeister, Christiane and Luca Benati**, “Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound,” *International Journal of Central Banking*, 2017, 9 (2).
- Bernanke, Ben**, “The Federal Reserve’s Balance Sheet: An Update,” 10 2010. Speech at the Federal Reserve Board Conference on Key Developments in Monetary Policy, Washington, D.C. [Accessed: 2020 09 21].
- , “The New Tools of Monetary Policy,” *American Economic Review*, 2020, 110(4), 943–983.
- Carpenter, Seth, Selva Demiralp, Jane Ihrig, and Elizabeth Klee**, “Analyzing Federal Reserve Asset Purchases: From Whom Does the Fed Buy?,” *Journal of Banking & Finance*, 2015, 52, 230–244.
- Cecchetti, Steve and Kermit Schoenholtz**, “The Brave New World of Monetary Policy Operations,” 2019. Blog post, available at <https://www.moneyandbanking.com/commentary/2019/6/14/the-brave-new-world-of-monetary-policy-operations>.
- Christensen, Jens and Glenn Rudebusch**, “The Response of Interest Rates to U.S. and U.K. Quantitative Easing,” *Economic Journal*, 2012, 122.
- and **Signe Krogstrup**, “Transmission of Quantitative Easing: The Role of Central Bank Reserves,” Working Paper 2014-18, Federal Reserve Bank of San Francisco, 2016.
- Ciccarelli, Matteo, Juan Angel Garcia, and Carlos Montes-Galdon**, “Unconventional Monetary Policy and the Anchoring of Inflation Expectations,” Working Paper Series 1995, European Central Bank, January 2017.
- Copeland, Adam, Antoine Martin, Gabriele La Spada, Marco Cipriani, Anna Kovner, and Gara Afonso**, “The Market Events of Mid-September 2019,” Staff Reports 918, Federal Reserve Bank of New York March 2020.
- ECB**, “Which Sectors Sold the Government Securities Purchased by the Eurosystem?,” Economic Bulletin Focus, European Central Bank, 2017.
- Ennis, Huberto**, “A Simple General Equilibrium model of Large Excess Reserves,” Working Paper WP 14-14, Federal Reserve Bank of Richmond, 2014.
- Eser, Fabian and Bernd Schwaab**, “Evaluating the Impact of Unconventional Monetary Policy Measures: Empirical Evidence from the ECB’s Securities Markets Programme,” *Journal of Financial Economics*, 2016, 119.
- Friedman, Milton and Anna J. Schwartz**, “Money and Business Cycles,” in Anna J. Schwartz, ed., *Money in Historical Perspective*, Chicago: University of Chicago Press, 1987, chapter 2, pp. 24–77.
- Gagnon, Joseph, Matthew Raskin, Julie Remache, and Brian Sack**, “The Financial Market Effects of the Federal Reserve’s Large-Scale Asset Purchases,” *International Journal of Central Banking*, 2011, 7 (1).

- Gambetti, Luca and Alberto Musso**, “The Macroeconomic Impact of the ECB’s Expanded Asset Purchase Programme (APP),” Working Paper Series 2075, European Central Bank 2017.
- Goodhart, Charles**, “Money, Credit and Bank Behaviour: Need for a New Approach,” *National Institute Economic Review*, 2010, *214.1*, F73–F82.
- , “The determination of the money supply: flexibility versus control,” *The Manchester School*, 2017, *85*, 33–56.
- Ihrig, Jane**, “Banks’ Demand for Reserves in the Face of Liquidity Regulations,” 2019. Blog post. Available at <https://www.stlouisfed.org/on-the-economy/2019/march/banks-demand-reserves-face-liquidity-regulations>.
- , **Ellen Meade, and Gretchen Weinbach**, “Monetary Policy 101: A Primer on the Fed’s Changing Approach to Policy Implementation,” Working Paper Series, FEDS 2015.
- Joyce, Michael, David Miles, Andrew Scott, and Dimitri Vayanos**, “Quantitative Easing and Unconventional Monetary Policy - An Introduction,” *The Economic Journal*, 2012, *564*.
- Kandrac, John and Bernd Schlusche**, “Quantitative Easing and Bank Risk Taking: Evidence from Lending,” *Journal of Money, Credit and Banking*, 2021.
- Keister, Todd and James Mc Andrews**, “Why Are Banks Holding So Many Excess Reserves?,” Staff Reports 380, Federal Reserve Bank of New York 2009.
- , **Antoine Martin, and James Mc Andrews**, “Divorcing Money from Monetary Policy,” *Federal Reserve Bank of New York Economic Policy Review*, 2008.
- Logan, Lorie**, “Observations on Implementing Monetary Policy in an Ample-Reserves Regime,” 2019. Remarks before the Money Marketeters of New York University, New York City. Available at <https://www.newyorkfed.org/newsevents/speeches/2019/log190417>.
- Ryan, Ellen and Karl Whelan**, “Quantitative Easing and the Hot Potato Effect: Evidence from Euro Area Banks,” *Journal of International Money and Finance*, 2021.

A Equilibrium proof

Monetary aggregates reach equilibrium levels once total repayments on other assets (Z_t) equal to new deposits (and other assets) created as a result of banks' unwanted reserves (O_t). As aggregates are constant over time in equilibrium, we derive equilibrium without time subscripts. Z can be expressed in terms of y , D and R as:

$$Z = \theta A = \theta(D - R) \quad (12)$$

As banks rebalance their asset after loan repayments are made, reserve outflows can be expressed

$$O = R - (1 - \beta)(D - W) \quad (13)$$

By construction aggregate loan repayments (Z) must equal deposit outflows caused by loan repayments (W) so we can sub our above expression for Z for W .

$$O = R - (1 - \beta)(D - \theta(D - R)) \quad (14)$$

Equilibrium values can then be derived by setting $O = Z$ and solving for the desired variable.

$$R - (1 - \beta)(D - \theta(D - R)) = \theta(D - R) \quad (15)$$

Equilibrium deposits are expressed as:

$$D = \frac{(1 + \theta\beta)R}{1 - \beta + \theta\beta} \quad (16)$$

B Equilibrium proof with capital

When capital is added to bank balance sheets $A + R = K + D$, where K is aggregate capital. Z can now be expressed as:

$$Z = \theta A = \theta(D + K - R) \quad (17)$$

And O as

$$O = R - (1 - \beta)(D - \theta(D + K - R)) \quad (18)$$

Equilibrium values can then be derived by setting $O = Z$ and solving for the desired variable.

$$R - (1 - \beta)(D - \theta(D + K - R)) = \theta(D + K - R) \quad (19)$$

Equilibrium deposits are now expressed as

$$D = \frac{(1 + \theta\beta)R - \theta\beta K}{1 - \beta + \theta\beta} \quad (20)$$

C Robustness testing and the role of market structure

It is difficult to exactly specify some of the model’s parameters using available data, namely the share of banks which ever receive reserves from the central bank (s), the share of banks which do so in each period (m), the share of assets sold to the central bank by non-banks (q) and the number of banks there should be in the model’s system (n). To ensure that our results in Section 5 are not primarily driven by incorrect specification of these parameters, we re-run our model across a range of plausible alternative values for each. Overall we find that β remains the central driving factor in determining the evolution of the money multiplier in our model and that conclusions drawn in Section 5 remain broadly unchanged. We also examine a number of implausible configurations to see how the market structure surrounding quantitative easing programmes may affect their transmission. When the injection of reserves into the financial system is more concentrated there is a smaller drop in the multiplier.

To illustrate the effects of the market structure around OMOs, we begin with two extreme and implausible calibrations. First, that all banks receive reserves in all periods ($m = s = 1$) and second that only one bank ever receives reserves from the central bank ($m = s = 0.01$). For simplicity we first run this model with β equal to 0.85 only. Figure 13a shows that these extreme assumptions create markedly different monetary dynamics, with more concentrated reserve injections *ceteris paribus* resulting in higher deposit growth and a smaller drop in the multiplier. This is because concentrated reserve injections will create unwanted reserves in individual banks earlier and in larger quantities. When β is high these parameters have less of an effect. In Figure 13b we show that with β of 0.99 they have virtually no effect. At this β the entire banking system enters reserve surplus as soon as reserves are increased, even when reserve growth is distributed evenly across banks.

Having illustrated the effects of these parameters in the extreme, we establish two plausible calibrations which could be used as alternatives to our baseline from the previous Section:

- **Alternative 1:** Only 60 per cent of banks ever receive reserves from the central bank ($s = 0.6$) and only 40 per cent of banks receive in each period ($m = 0.5$). This s value could capture the concentration of ultimate counterparty deposits (effectively institutional deposits) in certain parts of the banking system. By keeping m lower than s we continue to have a certain amount of randomness in a bank’s QE related reserve inflows each month.
- **Alternative 2:** We allow the whole system to receive reserves from the central bank at some point ($s = 1$), given how broad a set of QE counterparties are found in [Carpenter et al. \(2015\)](#). 90 per cent of these banks receive reserves from the central bank in each period ($m = 0.9$) so there is some randomness but not much.

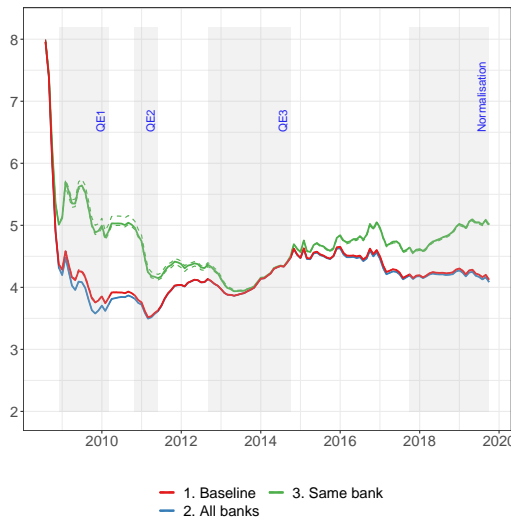
Figure 13c shows that these assumptions produce only marginal differences in model output.

We are also unsure of the exact share of assets sold to the Federal Reserve by non-banks but we do have a rough idea from [Carpenter et al. \(2015\)](#). In Figure 14a we examine the impact of marginal changes to our baseline figure of 90 per cent. Increasing this figure to 100 and decreasing to 80 has virtually no effect on model output. We also examine the effect of reducing this figure to zero, effectively showing how the money multiplier would have developed for an alternative programme geared entirely at the US banking system’s holdings of MBS and Treasuries. This produces a modestly lower money multiplier when aggregate reserves are increasing and higher when reserves are decreasing at the end. This is to be expected given the direct effect of central bank actions on deposit growth when non-banks are ultimate counterparties.

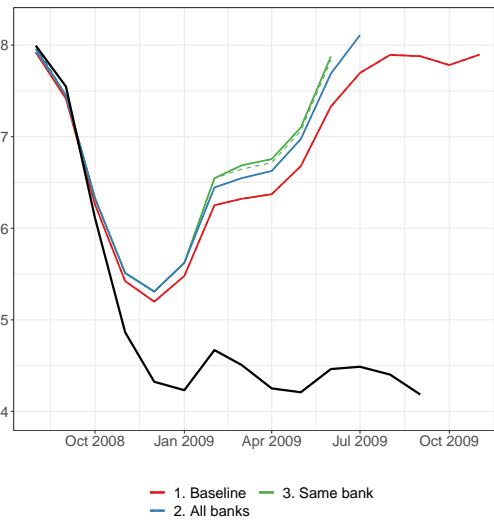
Finally, we run our model with 50, 100 (baseline), 200 and 400 banks in the system. Figure 14b shows that this produces essentially identical money multiplier dynamics.

Figure 13: Extreme assumptions regarding market structure can affect model outcomes at lower β values

(a) Money multiplier $\beta = 0.85$ and extreme market structure assumptions



(b) Money multiplier pre-QE β and extreme market structure assumptions



(c) Money multiplier with $\beta = 0.85$ and moderate changes in assumptions

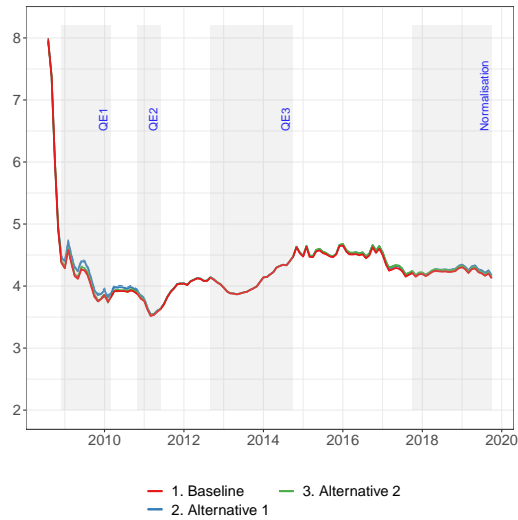


Figure 14: Other assumptions have limited implications for model outcomes

