Reserve Management and the Money Multiplier

Elia Moracci

Across the peak of the 2008 global financial crisis, the M1 money multiplier for the United States fell by about 40% in six months. An empirical analysis reveals that a surge in excess reserve holdings of depository institutions accounts for much of the observed variation in the multiplier. Attempting to explain this phenomenon as a shift in excess reserve holding policies of commercial banks, I formalize the problem they face exploiting inventory-theoretical models. I find that the introduction of large-sized negative jumps (used to model *bank runs*) in the law of motion for the excess reserve ratio produces discontinuous policy rules, replicating the observed variation better than in a standard model where the state evolves following exclusively continuous paths.

1 Introduction

From 2004 to 2018, the M1 money multiplier for the United States has fallen approximately from 1.6 to 0.9. This decrease hasn't occurred gradually: on the contrary, a sharp plunge took place from August 2008 to January 2009, at the peak of the 2008 financial crisis in the US, while the multiplier was relatively stable before and after this period.

I start from an empirical investigation of this phenomenon, trying to understand if an abrupt change in the aggregate behaviour of depository institutions or other agents caused the sudden variation in the multiplier. This analysis reveals that a 1000-billion-dollars surge in the monetary base, which wasn't matched by a corresponding increase in the M1 money stock, is responsible (from an accounting point of view) of the shift to the money multiplier. This mismatch between the evolution of the monetary base and of the M1 money stock was due to an increase in total reserves in the banking system (from 45 to 860 billion dollars in six months), whereas total deposits remained approximately stable. As reserve requirements didn't change during that short time-frame, the bulk of the increase in total reserves was due to a boost in excess reserves held by depository institutions. Until August 2008, the outstanding amount of excess reserves in the banking system was fluctuating around 2 billion dollars. Six months late, in January 2009, commercial banks in the US were holding an aggregate amount of about 800 billion dollars. Across the same period, the average excess reserve ratio in the US moved from 0.003 to 0.96, signalling a shift in the excess reserve holding behaviour of banks in the US.

I focus on this issue, trying to develop a behavioural theory of e, the excess reserve ratio. Building on a work by Craig and Koepke (2015), I describe the excess reserve holding problem faced by banks, highlighting the relevant trade-off they face as well as the main variables shaping their decision. Afterwards, exploiting empirical evidence which is coherent with previous treatments of this issue (Ennis and Wolman (2012), Berrospide (2013)), I describe how the determinants of banks' choices regarding excess reserves evolved during the period of interest, as a consequence of the deterioration of financial conditions and of the policy response implemented by the Federal Reserve.

I develop a theoretical framework in which I study optimal excess reserve management policies of commercial banks. The aim is to find a theoretical model that is able to reproduce the variation observed in the data as the consequence of a shift in the optimal policy triggered by changes in parameters, which represent the key determinants of excess reserve holding behaviour. An ideal framework to formalize the excess reserve holding problem faced by depository institutions is represented by *inventory-theoretical models*. These models are widely exploited for the analysis of optimal management of liquid assets. Economic agents, such as households or firms, need to employ cash or similarly liquid assets in order to perform their usual activities. The stock of liquid assets available in each moment for these agents is constantly influenced by physical operations (e.g., purchasing or selling goods) and by financial transactions (e.g., getting a loan, buying a security). The problem is to manage this stock of cash, choosing the optimal amount to hold at each point in time, given the following *trade-off*. On one hand, holding this liquid asset has an associate *opportunity cost*, which is given by the spread between the return on alternative investment opportunities and the return on the liquid asset (typically lower or even zero). On the other hand, holding a liquid asset delivers an essential benefit, which is the possibility of making a transaction whenever there is an opportunity to do so. Moreover, adjusting the amount of liquid asset held requires the payment of a *fixed cost*, which will have different interpretations in different models.

The use of inventory-theoretical models stems from the seminal work of Baumol (1952) and Tobin (1956), who analysed the problem of money demand by households in the presence of opportunity costs of holding cash and adjustment costs, when the consumption stream is completely *deterministic*, i.e.,

$$dm(t) = -\mu dt,$$

and the household does not receive cash from other sources than from its bank account, implying that cash balances never reach a level such that the household finds optimal to deposit some cash. Even though the Baumol-Tobin model is able to capture the fundamental determinants of the demand for liquid assets, its applicability is restricted by its deterministic nature. The work by Miller and Orr (1966) generalized the model to a setting where randomness plays a crucial role, the demand of money by firms. The possibility of random inflows and outflows of cash, modelled via the unregulated process

$$dm(t) = \sigma dW(t),$$

where $\{W(t)\}_{t\in\mathbb{R}_+}$ is a Wiener process, produces a feature not present in the Baumol-Tobin model, i.e., the coexistence of deposits and withdrawals. Several extensions of the Miller-Orr model have been studied in the last decades. Bar-ilan, Perry, and Stadje (2004) allow for a more general cost structure and for upward or downward jumps of random size in the stock. Alvarez and Lippi (2013) include the possibility of random negative jumps of size z, so that

$$dm(t) = \sigma dW(t) - z dN(t),$$

where $N(t)$ is the Poisson counter associated to a jump process with intensity κ.

I argue that excess reserves can be regarded as a liquid asset whose management is subject to the trade-off outlined above, since holding then has an associated opportunity cost and adjusting the stock held entails the payment of fixed costs. An application of inventory theory to the study of excess reserves management is due to Frost (1971), who proposed a model with both variable and fixed adjustment costs, obtaining a demand curve for excess reserves kinked at very low interest rates. The first contribution of this work is to exploit the setup proposed by Alvarez and Lippi (2013) to model excess reserve management by depository institutions, with the negative large random jumps representing bank runs. After describing formally the problem faced by commercial banks and the class of optimal control policies, I solve the model without bank runs ($\kappa = 0$) whose closed form analytical solution is useful to understand how the bank's choice concerning the optimal policy is affected by the various parameters of the model. Then, I go through the model with *bank run risk* ($\kappa > 0$), a non-trivial problem which requires a numerical solution. The second contribution of this work is the development of a routine that solves numerically the model with $\kappa > 0$ and $\sigma > 0$.[1]

After solving both models, I obtain implied aggregate statistics for an economy in which commercial banks follow the optimal control policy. These are the average excess reserve ratio M, the average number of adjustments per unit of time n and the stationary distribution of excess reserve ratios $h(m)$. I then review the quantitative properties of both models, in terms of elasticities of these aggregate statistics (and in particular of the average excess reserve ratio M) with respect to exogenous parameters. I find that the model with bank run risk produces *discontinuous optimal policies*, yielding stronger responses of the average excess reserve ratio to variations in parameters of the model than the model without bank runs, which is characterized by lower elasticities. This feature of the model with bank run risk suggests that this theoretical framework has the potential to explain the empirical evidence regarding the rise in excess reserve ratios of commercial banks, which future research may exploit.

This work is structured as follows. Section 2 introduces the money multiplier and document its fall during the 2008 financial crisis. Section 3 shows that the shift in the multiplier can be explained as a consequence of a massive increase in excess reserve holdings in US, and I describe the problem of excess reserve management faced by commercial banks. In Section 4 an inventory-theoretical model of excess reserve management is presented. The solution of the model without ($\kappa = 0$) and with ($\kappa > 0$) bank run risk is discussed in Section 5. Section 6 illustrates the results in terms of elasticities of average excess reserve ratios to exogenous parameters of the model. Section 7 concludes illustrating findings and limitations of the analysis and directions for future research about this issue.

[1] Alvarez and Lippi (2013) solved the model with $\kappa > 0$ and $\sigma = 0$, but they provided an algorithm to solve the more general model with $\sigma > 0$, which I exploited in this work.

2 The Fall of Money Multipliers in the US

A conventional view of the transmission of monetary policy rely heavily on the idea that the monetary authority can control indirectly the money supply by shifting the monetary base (Cukierman (2017)). Depending on the willingness of commercial banks to employ excess reserves for credit extension and on the propensity of the non-bank public to hold money in deposits rather than in cash, changes in the monetary base produce have smaller or bigger impacts on the money stock. To formalize this idea, an intuitive device has been developed and reached substantial popularity. This device is called the *money multiplier* and it tells how much a certain money stock m changes for a given change in the monetary base b[2], i.e.,

$$m = \mu \cdot b.$$

Hence, the money multiplier μ is commonly defined as a ratio of some money stock m to the monetary base b,

$$\mu = \frac{m}{b}. \tag{1}$$

For the theoretical formulation, following Mishkin (2016), I consider m as the sum of currency and deposits, that is

$$m = C + D.$$

Assuming that currency varies proportionally with deposits, I write

$$C = c \cdot D,$$

where $c > 0$ is called the *currency ratio*. I denote the monetary base b as the sum of currency and total reserves, i.e.

$$b = C + R + E,$$

where R is the amount of required reserves and E is the amount of excess reserves. Clearly, given the regulatory framework we will have

$$R = r \cdot D,$$

[2]For a description of monetary aggregates such as the M1 money stocks and the monetary base see Appendix A.2.

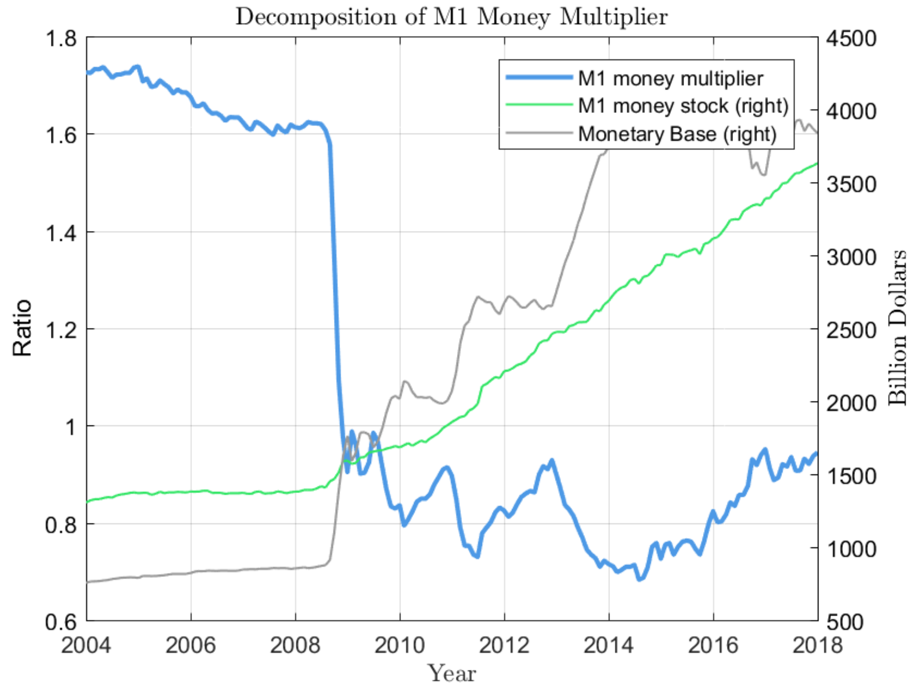


Figure 1: Evolution of M1 money multiplier and of its two components.

for some $r \in (0, 1)$ (the *required reserve ratio*), and

$$E = e \cdot D,$$

where $e \geq 0$ is the *excess reserve ratio*. We can then write the theoretical formulation of the money multiplier as

$$\mu = \frac{C + D}{C + R + E} = \frac{(1 + c)D}{(c + r + e)D} = \frac{1 + c}{c + r + e}. \tag{2}$$

In Figure 1, I displayed the time series plots for the M1 money multipliers for the United States in the period from January 2004 to January 2018. As the Figure shows, the M1 multiplier suddenly fell in the August 2008-January 2009 period, when it plunged from 1.6 to 0.9, a decrease of more than 40%. In the period from January 2004 to August 2008, the multiplier had a relatively stable trend, which was interrupted by the aforementioned fall, which occurred in the midst of the *2008 financial crisis*. From Figure 1 I can determine that, among the M1 money stock and the monetary base, the evolution of the latter had the larger impact on the multiplier: it experienced a massive boost from late 2008 to 2018, after a slow growth in previous years. In the last decade, it grew by around 3000 billion dollars, totalling an increase of almost 300%. More interestingly, the monetary

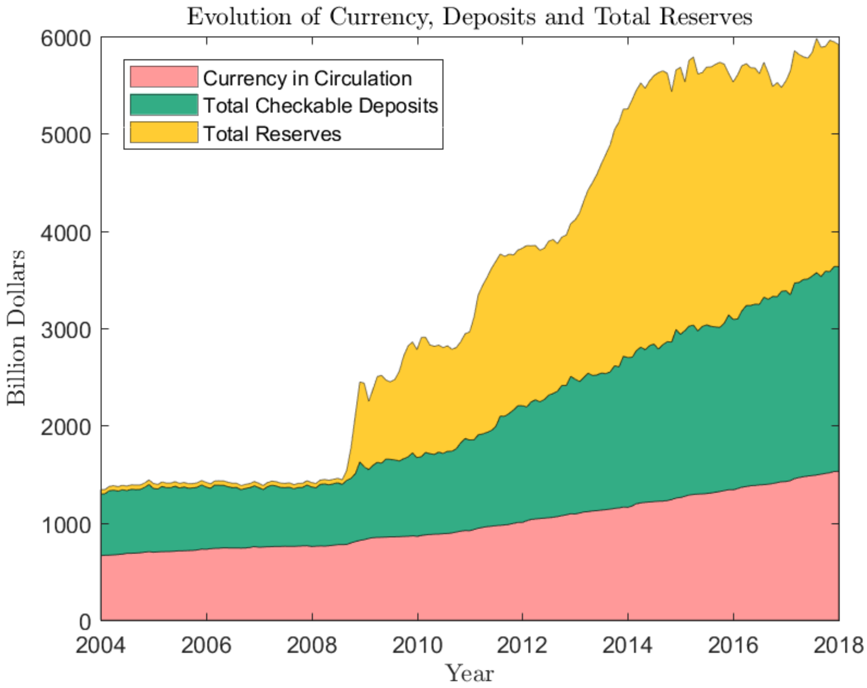


Figure 2: Evolution of the main components of monetary aggregates in the United States across the 2004-2018 period.

base displays a sudden shift across the peak of the financial crisis, growing by more than 100% in the period ranging from August 2008 to January 2009.

Figure 2 displays information about how the three subcomponents of M1 and MB (currency, total checkable deposits and total reserves) have changed over the period of analysis. This reveals that the observed increase in the monetary base, from an accounting point of view, is largely attributable to a surge in total reserves in the banking system, which grew from 45 billion dollars in August 2018 to 860 billion dollars in January 2009. The 1000-billion-dollars increase in the monetary base in this period is almost entirely attributable to the surge in reserves, as currency grew only by around 50 billions.

3 The Rise in Excess Reserve Holdings

In Section 2 I documented that the large plunge in the money multiplier that occurred in 2008 and lasted until today was driven by an increase in the monetary base and particularly in total reserves held by depository institutions. From Figure 3 it is clear that while required reserves grew steadily to keep complying with the regulatory framework (as total checkable deposits increased), a major change occurred in late 2008 concerning the excess reserve-holding behaviour of US depository institutions. As a matter of fact, ex-

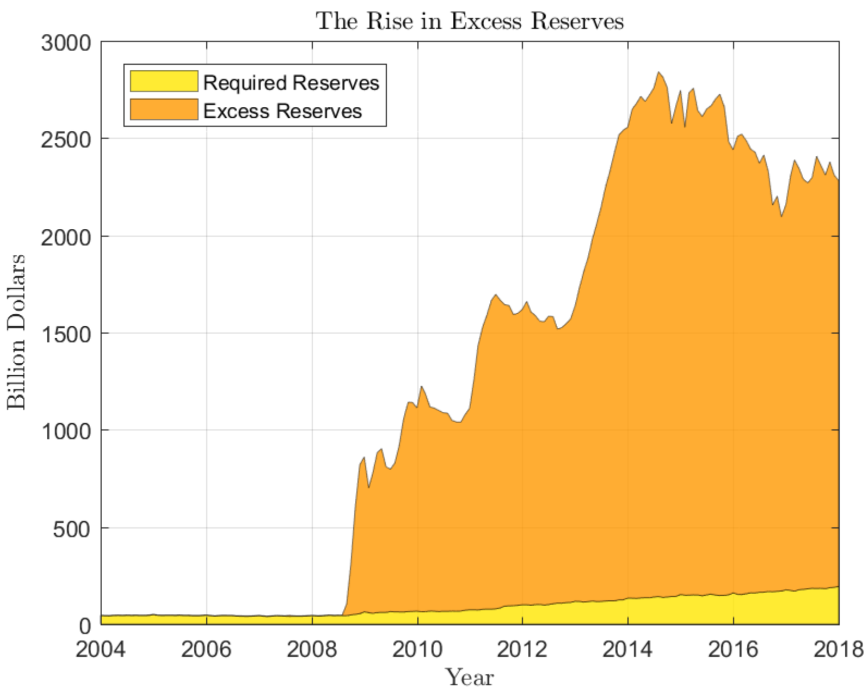


Figure 3: Evolution of required and excess reserves of depository institutions in the US from 2004 to 2018.

cess reserves of depository institutions in the US were negligible in August 2008, totalling less than two billion dollars. Starting from September 2008, they grew exponentially for a few months and at the beginning of the following year the outstanding amount was about 800 billion dollars. After this significant increase, excess reserves stationed around a trillion dollars until late 2010, where they started growing rapidly again reaching 1,5 trillion dollars. Another substantial increase can be observed starting from late 2012 until 2014. After that, reserve oscillated around a level of 2,5 trillion dollars until 2018. These movements in excess reserves explain the shift in the money multiplier documented in the previous Section. Recall the theoretical definition of the multiplier provided at page 6, i.e.,

$$\mu = \frac{C+D}{C+R+E} = \frac{(1+c)D}{(c+r+e)D} = \frac{1+c}{c+r+e}.$$

In Figure 4 I plot of e, r and c from 2004 to 2018. A significant increase in e took place at the peak of the crisis (in late 2008); instead, r was rather stable during all the period and c decreased only gradually from 2008 onwards. In Table 1 it is possible to observe the magnitude of the shift in the average excess reserve ratio in the US, which became about 350 times bigger in a matter of months.

In the remainder of this Section, I investigate the reasons behind these movements in

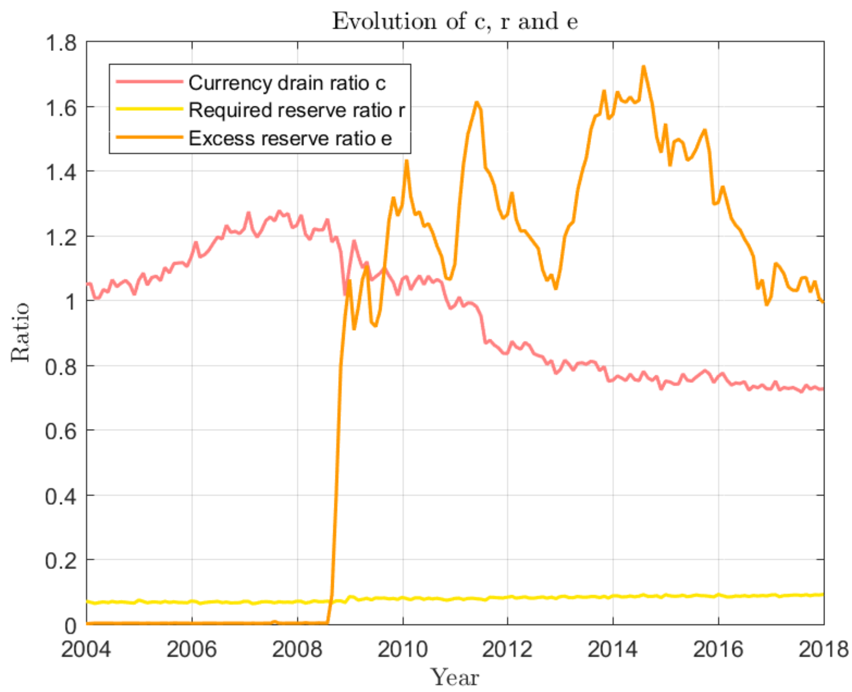


Figure 4: Evolution of e, r and c during the period of interest.

	August 2008	January 2009
e	**0.003**	**1.06**
r	0.071	0.084
c	1.25	1.10

Table 1: Evolution of e, r and c across the peak of the financial crisis in the US.

total excess reserves in the banking system. To understand the framework in which banks were taking decisions during the period of analysis, a basic knowledge of the *2008 global financial crisis* and of the policy response implemented by the FED in the US is required.

3.1 The 2008 Financial Crisis and FED's Policy Response

In this Subsection I provide a short summary of the 2008 financial crisis and I describe the policies implemented by the FED to improve economic conditions in the US. For the purpose of this thesis, I focus on how the crisis and the subsequent policy response altered the environment in which banks were taking their decisions. For a complete narrative of the financial crisis and of the related policy response, see Allen and Carletti (2010) and Minshkin and White (2014), which were the main reference for this summary.

After a steady increase in previous years, in late 2006 house prices in the US started plummeting. The collapse of the housing bubble led to mortgage delinquencies (refinancing loans became more difficult as adjustable-rates went up) which eventually resulted in the devaluation of mortgage-backed securities (MBSs), held by the majority of financial firms. As a consequence, in Summer 2007 interbank markets started experiencing pressures and borrowing conditions deteriorated. In late 2007 the price of subprime securitizations continued to fall and several financial institutions came under strains. After that Bear Stearns was bailed out by the FED in March 2008 through a merger with JP Morgan, on September 15, 2008 the investment bank Lehmann Brothers collapsed, increasing uncertainty and triggering panic on international financial markets. Signalling that credit risk in the financial industry might have been underestimated, Lehman's bankruptcy prompted investors to withdraw from the market and liquidity dried up. At this point, the interbank lending market went from being strained to being completely frozen.

From the first glimmers of the financial crisis in mid-2007 to 2013, the FED implemented several policies aimed at improve the functioning of interbank markets.

1. Starting from August 2007 and until January 2009 the FED's Board of Governors and the FOMC gradually reduced respectively the discount rate and the federal funds target rate, trying to ease borrowing conditions. Across the whole period both rates fell by around 5%. As Minshkin and White (2014) point out, this easy monetary policy was unusual in the first months, as the economy was still growing until December 2007.

2. The FED set up several facilities that provided liquidity, as the Term Auction Facility (TAF) in December 2007 and many other initiatives: the FED lending programmes reached a peak of 1.5 trillion dollars by the end of 2008.

3. The FED adopted a non conventional monetary policy of large scale asset purchases (LSAPs): in November 2008 it purchased 1.25 trillion dollars of mortgage-backed securities, to improve the condition of credit markets. This programme was labelled as QE1, the first round of *quantitative easing*, and it resulted in a massive expansion of the FED's balance sheet. Respectively in November 2010 and September 2012, the FED announced other two LSAP programmes, QE2 and QE3. In December 2013, the FED announced that the asset purchases would have been gradually phased

out over time.

4. On October 1, 2008, the FED started paying interest on required and excess reserve balances held by depository institutions. This rate was meant as an additional policy tool to maintain the effective federal funds rate close to its target. This policy action was particularly important for its impact on excess reserve holding decisions, as explained below.

As I discuss in Subsection 3.3, both the financial turmoil that originated from the crisis and these policy actions implemented by the FED had an impact on the excess-reserve holding behaviour of commercial banks.

3.2 The Excess Reserve Holding Problem

Commercial banks in the US hold reserves for two reasons. As briefly stated in the previous sections, the first reason is *compliance* with the regulatory framework: every depository institution must hold an amount of reserves (in cash or in its account at a Federal Reserve Bank) proportional to the amount of *deposit liabilities* it owes to its customers. The resources in the bank's vaults and accounts at FRBs that exceed this required amount are classified as *excess reserves*. Since there is no regulatory constraint that mandates depository institutions to hold excess reserves, their presence among banks' assets suggests that there is a second reason, the *willingness* to hold them. In particular, in deciding how much of this additional liquidity to hold, a financial institution should take into account two factors, as accurately described in Craig and Koepke (2015).

On one hand, since the return on excess reserves is smaller than the return on alternative risk-free investment opportunities, holding them has an associated opportunity cost, represented by the spread between the two interest rates. On the other hand, holding excess reserves can deliver an essential benefit to the bank: the possibility to meet its unknown needs for liquidity[3] and avoid paying the costs related with the lack of cash. Indeed, an unforeseen shortfall of liquidity may result in other types of costs: the impossibility to respond quickly to investment opportunities, make payments and service

[3] A depository institution has several sources of liquidity, not only reserves. Hence, zero excess reserves does not mean zero excess liquidity. However, when an outflow of liquidity takes place, both excess reserves and the other sources are typically eroded. Henceforth, excess reserves can be seen as a proxy for excess liquidity. In my simplified analysis, I assume that when the level of a bank's excess reserves falls below zero, that institution has a shortage of liquidity and should therefore immediately adjust its position, by borrowing funds on the interbank market or by divesting its asset holdings (sell Treasuries, cut lending, etc.).

deposit withdrawals, but also the interest rate that the bank will be charged in the interbank market to borrow liquidity and penalties imposed by the regulatory agency. The possibility of unforeseen insufficiency of liquidity is inevitably connected with the activities performed by depository institutions. Indeed, the fluctuations in their stock of liquid assets aren't determined solely by their actions, but also by external actors such as depositors. The impossibility to anticipate the variations in the stock of liquidity due to deposits and withdrawals exposes the bank to risk. An unusual concentration of withdrawals in a small amount of time can have a strong negative impact on the stock of liquidity held by the bank that may entail the costs described before. A purpose of holding excess reserves is to create a "buffer stock" that avoids this outcome.

A depository institution should henceforth choose the level of excess reserves that optimally balances the opportunity cost of holding them with its liquidity needs. The optimal level is influenced by several economic determinants: among these, three play a key role from the perspective of an optimizing agent.

- *Opportunity cost.* It is represented by the spread between the interest rate on excess reserves and that on alternative investment opportunities (loans, Treasuries, etc.). A higher spread should be associated with lower excess reserve holdings by depository institutions.

- *Adjustment costs.* The costs of adjusting the outstanding stock of excess reserves, which could be divided in

  - pure adjustment costs: costs that a bank must pay to adjust the stock, no matter the direction of the size of the adjustment (for example, paying an employee to lend out excess reserves or borrow more of them);

  - further costs related with forced upward adjustments (for example, penalties that should be paid to the regulatory agency and costs of borrowing the liquidity needed on the interbank market).

  Since unexpected outflows of liquidity that lead to forced upward adjustments are more likely when a bank holds few excess reserves, higher adjustment costs should induce a depository institution to hold more of them.

- *Volatility of bank balances.* The degree of variability in the stock of excess reserves due to inflows and outflows related with the daily operations of a bank, such as withdrawals and deposits. Higher volatility should induce banks to hold more excess reserves, to avoid paying the costs associated with a shortfall too often.

3.3 Determinants of Excess Reserves: Empirical Evidence

The increase in excess reserve holdings that occurred in late 2008 and the subsequent ones has been widely investigated in the literature and a number of reasons have been pointed out by different authors. However, several treatments of this issue argue that the above determinants of excess reserve holding patterns changed across the peak of the crisis, triggering a boom in the desired excess reserve holdings of commercial banks.

Craig and Koepke (2015) and Ennis and Wolman (2012) point out that the FED's policy response to the crisis had an impact on the opportunity cost of holding excess reserves, reducing it and thereby increasing their attractiveness. More specifically, policy actions by the FED and the FOMC altered in opposite directions the return on excess reserves and that on alternative investment opportunities, causing the opportunity cost of holding excess reserves to fall. Indeed, in the pre-crisis period, the return on excess reserves was zero, as the Federal Reserve hadn't implemented yet interest rate on excess reserves as a policy tool. From its enactment on October 1, 2008, the interest rate on excess reserves has fluctuated between 25 and 150 basis points. While the return on excess reserves went up, that on alternative investment opportunities plunged: for instance, the return on lending out resources (both on the interbank market and to households/firms) went down by about 5% from late 2007 to early 2009, as a consequence of the cuts to the federal funds target rate implemented by the FOMC. As Figure 5 shows, these two phenomena squeezed the spread, reducing the opportunity cost of holding reserves.

Other authors focus on the deterioration of financial conditions as a possible cause for liquidity hoarding by banks. Strains in the interbank lending market, together with increased uncertainty and concerns about solvency of financial institutions, could have had an impact both on adjustment costs and volatility of bank balances. In a situation of financial distress, both the probability that the level of excess reserves falls below zero and the adjustment cost that a bank must pay in this eventuality increase. As Ramos (1996) points out in his analysis of liquidity hoarding during the Great Depression in the 30s, during or immediately after a liquidity crisis, even a minor shortfall of cash may become a serious concern if depositors start doubting the institution's solvency. Thus, letting ex-

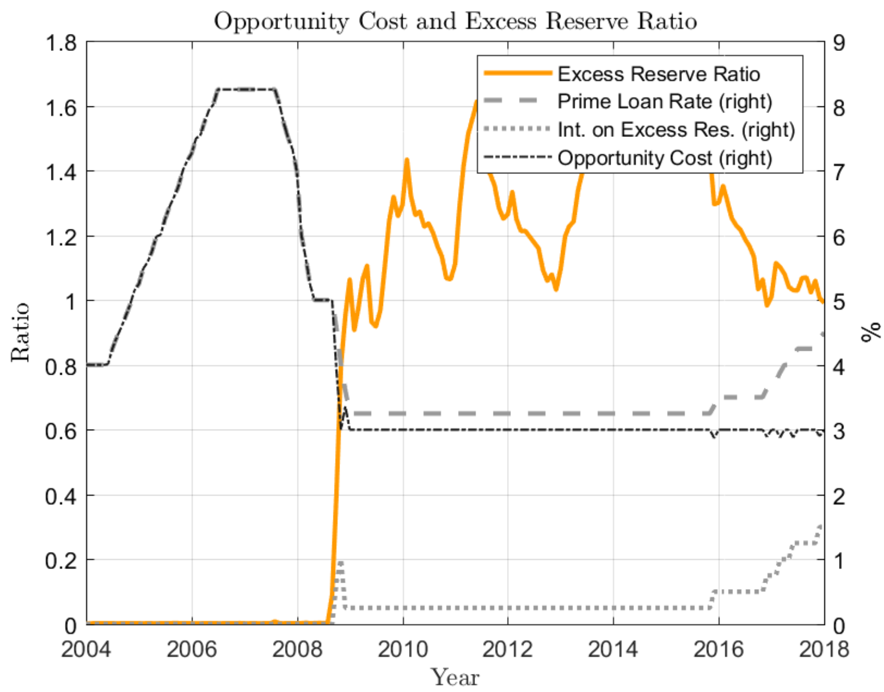


Figure 5: Excess reserve ratio e, prime loan rate, interest rate on excess reserves and spread between them from 2004 to 2018.

cess reserves fall below zero and being forced to borrow liquidity on the interbank market may inappropriately yield a bad signal that could trigger catastrophic consequences for the bank. To avoid paying this forced *adjustment cost* (a massive reputational cost) banks have an incentive to hold larger amount of excess reserves. Moreover, during a crisis lumpy cash outflows are more frequent due to widespread panic and uncertainty (*bank runs*). Therefore also the *volatility of bank balances* tends to be higher when financial markets come under strains. According to Berrospide (2013), the disruption in short-term funding markets that started in August 2007 and reached its peak in September 2008, together with falling asset prices and large withdrawals from institutional investors created a similar urge for the precautionary hoarding of liquidity. Unfortunately, it is difficult to obtain an empirical counterpart of adjustment costs and bank balances volatility. Regarding adjustment costs, possible proxies could be the effective federal funds rate or the discount rate, that can be used to estimate the cost of borrowing on interbank markets or from the Federal Reserve: however, this measure could represent only a tiny part of the perceived costs of borrowing (that may include very high reputational costs, as specified above). Concerning the volatility of bank balances, to construct an appropriate measure (as the one used by Frost (1971) in his empirical investigation) monthly bank-level data

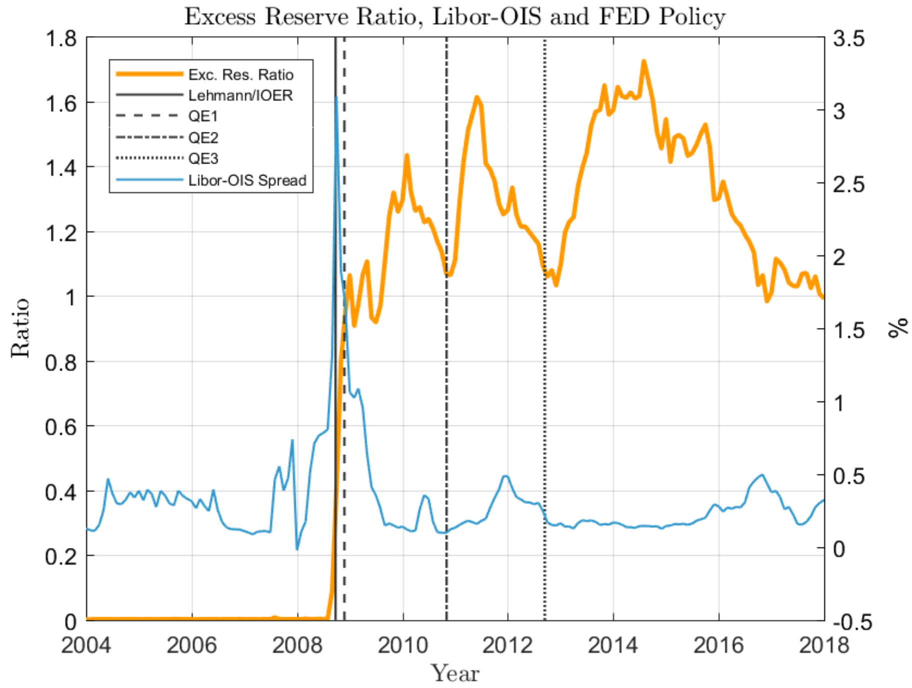


Figure 6: Excess reserve ratio e and Libor-OIS spread during the period of interest. The occurrence of five events is highlighted in the time series plot: the failure of the investment bank Lehmann Brothers/implementation of interest rate on excess reserves (which occurred within 20 days one from the other) and the announcements by the FED of the three rounds of quantitative easing (LSAPPs).

would be needed, in order to obtain figures for the quantity of deposits and withdrawals and derive the desired index. Without access to this kind of data, I must resort to a proxy, the *Libor-OIS spread*[4], a widely accepted measure of stress for the financial sector and illiquidity: since the latter two are associated with panic and larger withdrawals (as claimed by Ramos (1996) and Berrospide (2013)), the Libor-OIS spread could serve as an indicator for the volatility of bank balances. In Figure 6, I display the evolution of the excess reserve ratio e and the Libor-OIS spread. The Figure suggests that the peak in e observed in late 2008 might have been partly the result of increased uncertainty/illiquidity that boosted the volatility of bank balances, pushing banks to hoard liquidity.

This increased demand of excess reserves by financial institutions was matched by the FED's willingness to supply them. As summarized in Subsection 3.1, the FED started injecting liquidity into financial markets in December 2007: Walter and Courtois (2009) explain that in this initial phase, the injections took the form of loans extended to banks,

[4]The Libor-OIS spread is the difference between the LIBOR (an interest-rate average calculated from estimates submitted by the main banks in London where each bank estimates what it would be charged were it to borrow from other banks) and the OIS rate (the fixed rate of a overnight indexed swap, a type of interest rate swap). Additional details are provided in Sengupta and Tam (2008).

that provided as collateral securities that had devaluated due to the turmoil in financial markets. This action would have added reserves to the banking system: however, excess reserves didn't explode until September 2008. As Ennis and Wolman (2012) point out, this seeming inconsistency can be explained by the *sterilization* performed by the FED: since a surge in reserves could in principle push the fed funds rate below its target (threatening price stability), the central bank sold equal amounts of Treasuries to offset the effects of the injection on total excess reserves. Hence, until September 2008, total reserves in the system didn't change much. In October 2008 the FED started paying interest on reserves and stopped sterilizing the reserve creation due to FED's direct loans to depository institutions via the Term Auction Facility and the discount window; thereby excess reserves increased a lot. Starting from January 2009, the increases in excess reserves have been driven by asset purchases in the context of Large Scale Asset Purchase programmes (QE1, QE2, QE3). Figure 6 clearly highlights the relationship between policies implemented by the FED and waves of increases in the excess reserve ratio e.

4 Theoretical Framework

Consider a commercial bank which operates in a fractional-reserve banking system, so that it is forced to keep an amount of reserves in form of vault cash or as a credit balance in its account at the central bank. The bank has to choose optimally how many excess reserves to hold, facing a trade-off: holding excess reserves has an associated cost, which is measured by the spread between the return given by investing the money in an alternative asset (e.g., a loan or a security) and the payment received from the central bank on excess reserve holdings. At the same time, each upward or downward adjustment of the stock requires the payment of a fixed cost. The sequence problem for the commercial bank is given by

$$V(m) = \min_{\{m(t),\tau_i\}} \mathbb{E}\left(\int_0^{+\infty} e^{-rt} Rm(t)dt + \sum_{i=0}^{\infty} e^{-r\tau_i} b \,\middle|\, m(0) = m \right),$$

subject to

$$dm(t) = \sigma dW(t) - zdN(t), \tag{3}$$

where m is the bank's excess reserve ratio $e = E/D$[5], R is the opportunity cost, b is the adjustment cost, $\{\tau_i\}_{i\in\mathbb{N}}$ are the *stopping times* at which adjustments occur (defined via a *stopping rule*, which I discuss below), $\{W(t)\}_{t\in\mathbb{R}_+}$ is a Wiener process and $\{N(t)\}_{t\in\mathbb{R}_+}$ is the Poisson counter associated with a jump process with rate of arrival κ.

The above model, based on Miller and Orr (1966) and Alvarez and Lippi (2013), seems particularly adequate to model the environment in which banks operate due to its inclusion of both small and frequent random inflows and outflows of cash (via the generalized Brownian motion) and large random outflows of cash that represent the possibility of being subject to a *bank run* (via the jump process with size z and intensity κ). Moreover, the three determinants of excess reserve holdings mentioned in Subsection 3.1 match three parameters of the model: the opportunity cost is represented by the parameter R, the adjustment cost is represented by the parameter b and the volatility of bank balances is represented by the parameter σ.

The bank must choose an *impulse-control policy*, a rule that determines if and how to adjust the level of excess reserves given the current observable level. I consider *impulse-control band* (ICB) policies, characterized by a *trigger* level $\bar{m}$ and a *target* level m^* (with $m^* < \bar{m}$). The bank will let the excess reserve ratio m fluctuate freely inside the region $(0, \bar{m})$ (the *inaction region*, whose lower limit is imposed by the regulatory framework as the excess reserve ratio cannot fall below zero) given the law of motion (3), until they reach 0 or $\bar{m}$. Reaching a threshold will immediately trigger an instantaneous adjustment to the target level m^*. Figure 7 portrays a control policy of the ICB type. The *stopping rule* implied by this policy that specifies *stopping times* $\{\tau_i\}_{i\in\mathbb{N}}$ is

$$\tau_0 = \min\{t | m(t) \leq 0 \vee m(t) \geq \bar{m}\}, \quad \text{and} \quad \tau_i = \min\{t > \tau_{i-1} | m(t) \leq 0 \vee m(t) \geq \bar{m}\}.$$

Then, the problem becomes

$$V(m) = \min_{\{m^*, \bar{m}\}} \mathbb{E}\left(\int_0^{+\infty} e^{-rt} R m(t) dt + \sum_{i=0}^{\infty} e^{-r\tau_i(m^*,\bar{m})} b \,\middle|\, m(0) = m \right), \tag{4}$$

since the bank choice is reduced to the specification of an ICB policy $(m^*, \bar{m})$. As discussed in Alvarez and Lippi (2013) (Section 3.4), in this context ICB policies are the opti-

[5]The excess reserve ratio $e = E/D$ is more appropriate that the absolute level of excess reserves E, since the latter alone is not informative. Indeed, to determine if a certain level of excess reserves is high or low, it should be compared to the outstanding stock of deposit liabilities of the bank.

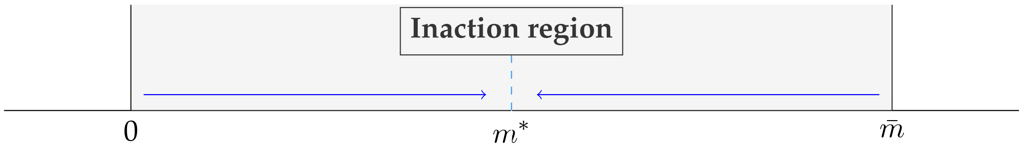


Figure 7: The impulse-control band class of policies. In the inaction region the process is not subject to any control, but when it reaches the trigger levels 0 and $\bar{m}$ control is exercised and the process is reverted back to m^*, the target level, with a lumpy adjustment that requires the payment of a fixed cost.

mal control policies.

5 Solving the Model

I now present the solution of the above model both without bank runs (whose tractability allows for an analytical characterization of optimal policy rules) and with bank run risk.

5.1 The Model without Bank Runs

Here, I discuss the solution of the model without bank runs, characterized by $\kappa = 0$. I also let $r = 0$ (no discounting, all periods are equally weighted). These simplifications simplify the model enough to make it analytically tractable, managing to obtain closed form solutions that yield significant insights regarding the effect of parameters of the model on the optimal control policy chosen by the commercial bank. Without discounting, the problem is a so-called *steady state* problem, where the commercial bank solves

$$\min_{M,n} RM + bn, \tag{5}$$

where M is the average excess reserve holding and n is the average number of adjustments per unit of time. The higher will be the number of adjustments n, the lower will be the average reserve holding M, and vice-versa I rewrite the cost minimization problem (5) in terms of ICB policy $(m^*, \bar{m})$, writing

$$\min_{m^*,\bar{m}} RM(m^*, \bar{m}) + bn(m^*, \bar{m}), \tag{6}$$

since average reserve holding M and the average number of adjustments per unit of time n are dependent on the chosen upper trigger level $\bar{m}$ and on the target level m^*. In order to solve (6), we need to find the functions $M(m^*, \bar{m})$ and $n(m^*, \bar{m})$. To derive the average number of adjustments per unit of time, I first derive the average time between two adjustments (*expected length of a cycle*) $T(m^*, \bar{m})$ and then I compute $n(m^*, \bar{m}) = 1/T(m^*, \bar{m})$.

Proposition 1 *Let $\kappa = 0$ and $r = 0$. The average number of adjustments per unit of time is then given by*

$$n(m^*, \bar{m}) = \frac{1}{T(m^*.\bar{m})} = \frac{\sigma^2}{m^*(\bar{m} - m^*)}. \tag{7}$$

Proof. See Appendix B.1. ■

Corresponding to intuition, the average number of adjustments per unit of time is increasing in volatility σ^2: higher variability of the unregulated process implies that the barriers in expectation are reached more often. To determine the average excess reserve holdings $M(m^*, \bar{m})$, consider the invariant distribution characterized by

$$h(m) = h(m, t), \quad \text{for } m \in [0, \bar{m}] \quad \forall t \in \mathbb{R}_+,$$

so that the probability that the excess reserve ratio is at some specific level is independent of time t. Integrating over the invariant distribution of excess reserve ratios, I can determine $M(m^*, \bar{m})$.

Proposition 2 *Let $\kappa = 0$ and $r = 0$. The invariant distribution of the excess reserve ratio is then given by the triangular distribution*

$$\begin{cases} h(0) = h(\bar{m}) = 0 & \\ h(m) = \dfrac{2m}{m^*\bar{m}}, & \textit{for} \quad m \in (0, m^*], \\ h(m) = \dfrac{2}{(\bar{m} - m^*)} - \dfrac{2m}{\bar{m}\,(\bar{m} - m^*)}, & \textit{for} \quad m \in (m^*, \bar{m}). \end{cases}$$

The average excess reserve holdings are given by

$$M(m^*, \bar{m}) = \mathbb{E}\,(m) = \int_0^{\bar{m}} m h(m) dm = \frac{\bar{m} + m^*}{3}. \tag{8}$$

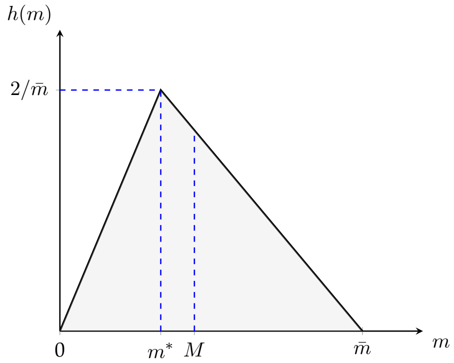


Figure 8: The stationary probability distribution of excess reserve holdings, which is a triangular probability distribution. M is the mean of the distribution, while m^* is the mode.

Proof. See Appendix B.2. ■

The invariant distribution of excess reserve ratios and the average excess reserve ratio are displayed in Figure 8. Having expressed M and N as functions of m^* and $\bar{m}$, it is possible to rewrite the problem in a more convenient way

$$V(m,b) = \min_{m^*,\bar{m}} \frac{R\left(\bar{m}+m^*\right)}{3} + \frac{b\sigma^2}{m^*(\bar{m}-m^*)}.$$

Proposition 3 *Let $\kappa = 0$ and $r = 0$. The optimal target value m^* and the optimal upper trigger value $\bar{m}$ are then given by*

$$m^* = \left(\frac{3b\sigma^2}{4R}\right)^{1/3}, \quad \bar{m} = 3m^* = 3\left(\frac{3b\sigma^2}{4R}\right)^{1/3}. \tag{9}$$

Proof. Straightforward optimization delivers the desired result. ■

5.2 The Model with Bank Run Risk

I now move to the model with bank run risk, characterized by $\kappa > 0$. I also allow for discounting of future periods by the commercial bank ($r > 0$) and I represent the sequence problem (4) in a recursive form using a discrete-time discrete-state approximation to derive the following *Hamilton-Jacobi-Bellman equations*, which can be used to pin down the optimal values m^* and $\bar{m}$[6]. If $m \leq z$, we have

$$V(m) = \min\left\{b + \min_{\hat{m}} V(\hat{m}), \frac{Rm + \frac{V''(m)\sigma^2}{2} + \kappa\left(b + \min_{\hat{m}} V(\hat{m})\right)}{r+\kappa}\right\}.$$

whereas if $m > z$, we have

$$V(m) = \min\left\{b + \min_{\hat{m}} V(\hat{m}), \frac{Rm + \frac{V''(m)\sigma^2}{2} + \kappa \min\left\{b + \min_{\hat{m}} V(\hat{m}), V(m-z)\right\}}{r+\kappa}\right\}.$$

At each point in time the commercial bank has two choices. It can adjust, by paying a fixed amount b and reducing/increasing its excess reserve ratio to the optimal one. Alternatively, it can wait, facing the flow cost and the expected future costs given the underlying process for m. According to the value of m with respect to z, the bank may have different possibilities. If $m > z$, indeed, in the case of a run the bank is able to decide whether to adjust or not. If $m \leq z$, it is forced to adjust. I define the optimal return point as the minimal argument of the value function, i.e.,

$$m^* = \arg\min_{\hat{m}} V(\hat{m}). \tag{10}$$

In the interior of the inaction region, i.e., for $0 < m < \bar{m}$,

$$rV(m) = Rm + \frac{V''(m)\sigma^2}{2} + \kappa\left(V(m^*) + b - V(m)\right) \quad \text{for } 0 < m \leq z < \bar{m}, \tag{11a}$$
$$rV(m) = Rm + \frac{V''(m)\sigma^2}{2} + \kappa\left(V(m-z) - V(m)\right), \quad \text{for } 0 < z < m < \bar{m}. \tag{11b}$$

Equation 11a is second order non-homogeneous ordinary differential equation, while Equation 11b is a second order non-homogeneous *delay differential equation*. To emphasize the fact that $V(m^*)$ is a constant, we denote it by V^* from now on. The problem has several accompanying boundary conditions. First, by definition of m^*, if V is differen-

[6]Following Dixit (1993), for a discounted expected value $F(m) = \mathbb{E}\left[\int_0^\infty e^{-\rho t} f\left(m(t)\right) | m(0) = m\right]$, splitting the integral into $(0, \Delta]$ and $(\Delta, +\infty)$, in absence of adjustment of m we have

$$F(m) = f(m)\,\Delta + \frac{1}{1+\rho\Delta}\mathbb{E}\left(F(m+\Delta m)\right).$$

In our case $F(m) = V(m), f(m) = Rm$ and $m(t)$ follows (3), using Ito's Lemma (see Stokey (2008)), dividing by Δ and taking $\Delta \to 0$ we can derive

$$(\rho+\kappa)V(m) = Rm + \frac{\sigma^2 V''(m)}{2} + \kappa\left(V(m-z)\right).$$

tiable we must have

$$V'(m^*) = 0. \tag{12}$$

The *value matching* conditions for this problem are

$$V(m) = V(m^*) + b = V^* + b, \quad \text{for } m \leq 0, \tag{13a}$$

$$V(\bar{m}) = V(m^*) + b = V^* + b, \quad \text{for } m \geq \bar{m}. \tag{13b}$$

If V is differentiable at $m = \bar{m}$ we also have the *smooth pasting* condition

$$V'(\bar{m}) = 0. \tag{14}$$

The following Proposition describes the function V that solves (11a) and (11b) complying with the associated boundary conditions, given a specified policy $(m^*, \bar{m})$.

Proposition 4 *Given a policy specified by thresholds $0 < m^* < \bar{m}$, the value of following such a policy can be described by J functions V_j*

$$V(m, m^*, \bar{m}) = V_j(m), \ \textit{ for } m \in [zj, \min\{z(j+1), \bar{m}\}],$$

where

$$V_j(m) = A_j + D_j(m - zj) + \sum_{k=1}^{2} \sum_{i=0}^{j} B_{j,i}^k e^{\lambda_k (m - zj)} \cdot (m - zj)^i,$$

where $\lambda_1 > 0 > \lambda_2$ are the solutions of $r + \kappa = (\sigma^2/2)\lambda^2$ and the constants A_j, D_j, $B_{j,i}^k$ for $j = 1, ..., J-1$, $i = 0, 1, ..., j$ and $k = 2$ solve a block recursive system of linear equations described in the proof.

Proof. See Appendix B.3. ∎

The constants A_j, D_j, $B_{j,i}^k$ are functions of the optimal policy $(m^*, \bar{m})$, which should be pinned down using *optimality of return point* (Equation 12) and *smooth pasting* (Equation 14). Applying these two conditions to the newly derived function V, we obtain a system

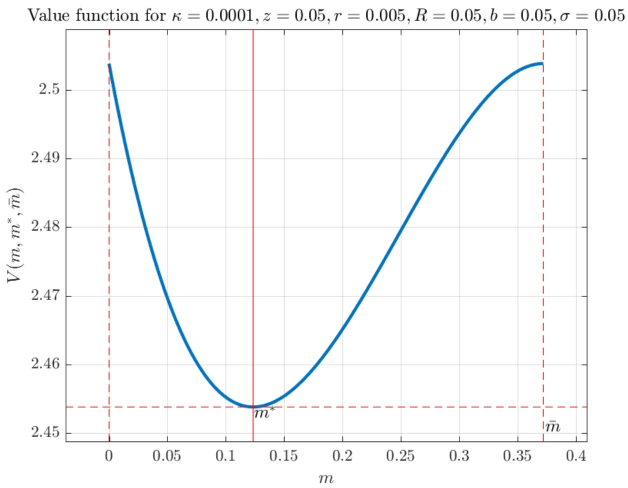


Figure 9: Value function $V(m, m^*, \bar{m})$ where $(m^*, \bar{m})$ is the optimal policy for displayed parameter values.

of two equations determining $(m^*, \bar{m})$.

$$0 = V'(m^*) = D_{j^*} + \sum_{k=1}^{2} \sum_{i=0}^{j^*} B^k_{j^*,i} e^{\lambda_k (m^* - zj^*)} \left(\lambda_k \left(m^* - zj^*\right)^i + i \left(m^* - zj^*\right)^{i-1} \right),$$

$$0 = V'(\bar{m}) = D_{J-1} + \sum_{k=1}^{2} \sum_{i=0}^{J-1} B^k_{J-1,i} e^{\lambda_k (\bar{m} - z(J-1))} \left(\lambda_k \left(\bar{m} - z(J-1)\right)^i + i \left(\bar{m} - z(J-1)\right)^{i-1} \right).$$

As discussed in Alvarez and Lippi (2013) for the case $\sigma = 0$, the nature of the above boundary conditions is necessary, not sufficient. I avoid to make statements about the nature of boundary conditions for the case $\sigma \geq 0$. I exploit a Matlab routine to find the optimal policy $(m^*, \bar{m})$ that yields the smallest $V^*(m^*, \bar{m})$. The Matlab program that solves this problem is described in Appendix C.1. In Figure 9 I plot the value function $V(m, m^*, \bar{m})$, where $(m^*, \bar{m})$ is the optimal policy for given parameter values. Notice that in this case the optimal return point is big enough to cover for a jump since $m^* > z$: this happens as long as $m > z$, after which the agent enters in a *danger zone* in which any jump triggers a forced adjustment.

6 Results

Having obtained the optimal policy $(m^*, \bar{m})$, it is possible to retrieve three key aggregate statistics for an economy in which a continuum of commercial banks follow the optimal control policy: the average excess reserve ratio $M(m^*, \bar{m})$, the average number of adjustments $n(m^*, \bar{m})$ and the stationary distribution of excess reserve ratios $h(m, m^*, \bar{m})$. If the problem is tractable (as for the case $\kappa = 0$) the statistics can be derived analytically; otherwise, one may resort to numerical techniques or to computer simulations of the controlled process. To determine how well this theoretical framework reproduces the shift in the observed aggregate excess reserve ratio as the result of a change in some underlying determinant of commercial banks' behaviour, I perform comparative statics of M and n with respect to parameters of the model.

6.1 Results for the Model with $\kappa = 0$

Proposition 5 *Let $\kappa = 0$ and $r = 0$. The average excess reserve ratio M and the average per-period number of adjustments n are*

$$M = \frac{4}{3}\left(\frac{3b\sigma^2}{4R}\right)^{1/3}, \quad \textit{and} \quad n = \frac{\sigma^2}{2}\left(\frac{3b\sigma^2}{4R}\right)^{-2/3}.$$

Proof. Substituting the values found for m^* and $\bar{m}$ in Equation 7 and Equation 8, the result follows. ■

One can perform some basic *comparative statics*, in order to get some insights on the evolution of the optimal policy (and consequently of the relevant statistics) when some exogenous parameter changes. Taking partial derivatives we obtain

$$\frac{\partial M}{\partial b} > 0, \quad \frac{\partial M}{\partial \sigma} > 0, \quad \frac{\partial M}{\partial R} < 0, \quad \frac{\partial n}{\partial b} < 0, \frac{\partial n}{\partial \sigma} < 0, \text{ and } \frac{\partial n}{\partial R} > 0.$$

An increase in the cost of adjustment b, increasing both the target and the upper trigger, will expand the inaction set, thereby leading to a smaller number of average adjustments per unit of time and to a larger average excess ratio. This is intuitive: as the cost of a lumpy adjustment rises, the bank will choose to exercise control less frequently, keeping more excess reserves on average. An increase in volatility σ will have the same effect: higher volatility increases the fluctuations in the unregulated process for the excess reserve ratio. Hence, the optimal inaction set is expanded, having the same effects on M and N as b. Conversely, an increase in the opportunity cost R will lower both the target and the trigger levels, causing a reduction in the average excess reserve ratio M and an increase in the number of adjustments N, corresponding to intuition.

I compute elasticities to evaluate the quantitative implications of the model: since the

	August 2008	December 2008	Increase	% Increase
e	0.003	0.94	313,3 times bigger	31233%

Table 2: The actual increase in e at the peak of the 2008 financial crisis.

aim is to explain the rise in the excess reserve ratio, I focus on the elasticities of M with respect to the parameters of the model, i.e.,

$$\frac{\partial \log M}{\partial \log b} = \frac{1}{3}, \quad \frac{\partial \log M}{\partial \log \sigma} = \frac{2}{3}, \quad \text{and} \quad \frac{\partial \log M}{\partial \log R} = -\frac{1}{3}.$$

Even if the qualitative implications of the model correspond to economic intuition, it performs very badly from a quantitative point of view. In Table 2 I display the observed variation in the excess reserve ratio at the peak of the financial crisis. It is immediate to notice that with the elasticities displayed above, massive variations of the parameters of the model in a short period of time are needed to explain the observed sharp rise in e.

6.2 Results for the Model with $\kappa > 0$

The computation of the statistics $M(m^*, \bar{m}), n(m^*, \bar{m})$ and $h(m, m^*, \bar{m})$ for the model with bank run risk is more involved[7]. Starting from M and n, I define

$$M(m) = \mathbb{E}\left[\rho \int_{t=0}^{\infty} e^{-\rho t} m(t) \Big| m(0) = m\right],$$
$$n(m) = \mathbb{E}\left[\rho \sum_{j=0}^{\infty} e^{-\rho \tau_j} \Big| m(0) = m\right],$$

respectively the expected discounted integral (at rate ρ) of an adjustment indicator (τ_j are the stopping times at which adjustments occur) and of cash balances, both conditional on the current value of m. Expected values are taken with respect to the law of motion of m.

Proposition 6 *The aggregate statistics M and n are given by*

$$M = \lim_{\rho \to 0} M(m), \quad \textit{and} \quad n = \lim_{\rho \to 0} n(m),$$

where the functions $n(m)$ and $M(m)$ can be found solving an ODE-DDE system described in the

[7]The Matlab routine I exploited to solve for M, n and $h(m)$ is described in Appendix C.2.

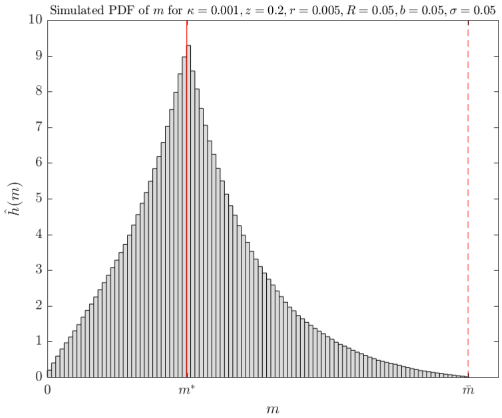


Figure 10: Approximated stationary distribution of excess reserve ratios $h(m, m^*, \bar{m})$ where $(m^*, \bar{m})$ is the optimal policy for displayed parameter values.

proof.

Proof. See Appendix B.4. ∎

To obtain the distribution of excess reserve ratios $h(m, m^*, \bar{m})$, I simulated a Brownian motion with jumps subject to the optimal control policy $(m^*, \bar{m})$ and derived an approximated counterpart of $h(m, m^*, \bar{m})$, which I display in Figure 10.

I now perform *comparative statics*, determining how policy rules $(m^*, \bar{m})$ and implied aggregate statistics M and n change as κ and R vary.

- The first two plots in Figure 11 display comparative statics of policy rules and aggregate statistics with respect to κ. There exist a threshold $\bar{\kappa}$ such that if $\kappa \geq \bar{\kappa}$ a jump in m^* realizes: for a high enough rate of arrival of bank runs κ, commercial banks find it optimal to switch to a policy characterized by higher average excess reserve ratios. Since *discontinuous policy rules* arise there is a jump in the aggregate statistics M and n for $\kappa > \bar{\kappa}$. Notice that across the threshold $\bar{\kappa}$ the elasticity of M to κ is degenerate. The existence of piecewise-defined non-differentiable policy rules is a consequence of the introduction of large-sized exogenous variations in m, as I show in Appendix B.5. In this stochastic model, however, optimal policies are not only non-differentiable but also discontinuous. As the value function is differentiable, discontinuity of optimal policies seems to be in contrast with a well-known result in dynamic programming: differentiability of the value function implies continuity of the policy function (see Stokey, Prescott, and Lucas (1989)). In fact, the lack of convexity in the *feasibility set* in this framework, which is an assumption

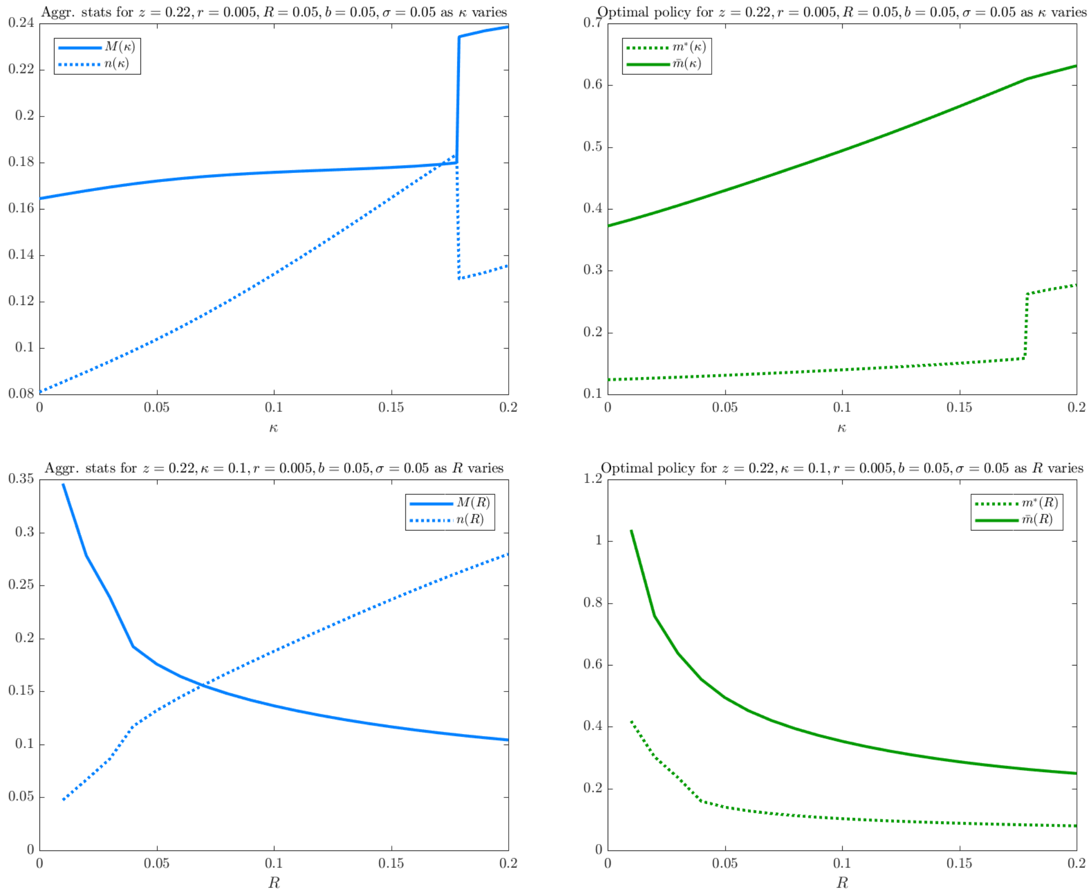


Figure 11: Comparative statics of the optimal policy $(m^*, \bar{m})$ and of implied aggregate statistics $M(m^*, \bar{m})$ and $n(m^*, \bar{m})$ with respect to κ and R.

needed to obtain the above result, is responsible for this apparent inconsistency. In this model, non-convexities in possible choices stem from the presence of fixed costs. For instance, suppose that $m = 100$ and $b = 1$: the agent can withdraw 99 paying 1 and can withdraw $98 = 49 + 49$ paying 2 (two withdrawals), but he cannot withdraw $98, 5$ paying $1, 5$. Hence, there are linear combinations of feasible actions which are not feasible, implying that the feasibility set is not convex. Thus, policy rules may be discontinuous. To understand why jumps occur in this framework for some parameter values, observe the shape of value function in Figure 12 where $\kappa = \bar{\kappa}$. It is possible to see that for this specific value of κ, the value function has two global minima, i.e., for the commercial bank there is indifference between these two excess reserve ratios. For $\kappa \neq \bar{\kappa}$ indifference is resolved: for $\kappa < \bar{\kappa}$ the only global minimum would be the smaller one, and vice-versa. Henceforth, tiny variations in κ may trigger jumps in m^* and also in the aggregate statistics.

- The last two plots of Figure 11 display comparative statics with respect to R. As

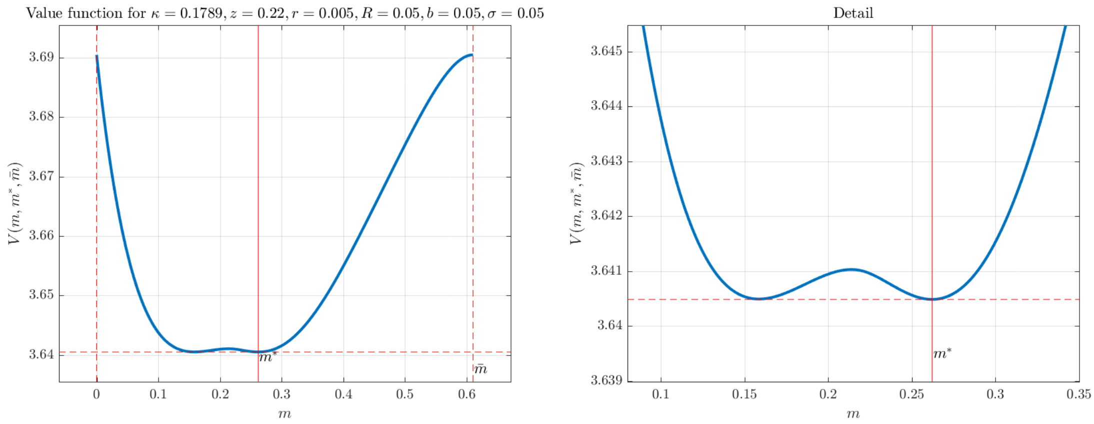


Figure 12: Value function $V(m, m^*, \bar{m})$ for $\kappa = \bar{\kappa}$. The detail highlights the presence of indifference between two global minima. For $\kappa < \bar{\kappa}$ and $\kappa > \bar{\kappa}$ indifference is resolved.

already seen in the $\kappa = 0$ case, the average excess reserve ratio M decreases with the opportunity cost since the inaction region shrinks when R rises, implying that adjustments are more frequent. However, in that case the elasticity was constant for all R and equal to $-1/3$. In this case, for low and high values of R the elasticity is still approximately $-1/3$, but there is a region (in the plot, approximately for $R \in (0.03, 0.04)$) in which the elasticity is much higher in magnitude (around $-4/3$). This feature suggests that for average values of the spread R, banks are more sensitive to its changes in deciding optimal excess reserve holdings that when R is very low or very high.

7 Concluding Remarks

I started from an empirical analysis of the evolution of the money multiplier in the United States in the last fifteen years, trying to determine the reasons behind its plunge in late 2008. I found that a substantial increase in excess reserve holdings of depository institutions was the main factor causing that shift and I thereby focused on the excess reserve holding problem faced by commercial banks, trying to develop a behavioural theory of e (the ratio between excess reserves and deposits, a measure of excess liquidity) that could explain its rise as the consequence of variations in the determinants of optimal excess reserve management policies of commercial banks.

To formalize the issue, I presented an inventory-theoretical model of excess reserve management that exploits the setup proposed by Miller and Orr (1966) and Alvarez and

Lippi (2013). This model includes an exogenous process for the excess reserve ratio that has two components: a continuous one (modelled via a Brownian motion without drift, that represents the small and frequent fluctuations in the excess reserve ratio due to frequent small-sized deposits and withdrawals) and a discontinuous one (modelled via random large-sized negative jumps which occur following a Poisson process, that represents infrequent bank runs).

I find that the inclusion of negative jumps produces discontinuous policy rules, generating much higher elasticities of the excess reserve ratio with respect to parameters of the model compared to a framework where the excess reserve ratio evolves following exclusively continuous paths (no possibility of bank runs). I show that the discontinuity in optimal policy is motivated by the existence of multiple local minima for certain parameter values and by the fact that a small variation in the probability of bank runs κ may cause a shift in the optimal policy from a minimum to another one, causing a jump in the average quantity of excess reserves held by commercial banks and a decrease in the frequency of adjustments. I also find that for average values of R the elasticity of the average excess reserve ratio with respect to the opportunity cost R is higher than the one implied by a standard Miller-Orr model.

The observed discontinuity in optimal policies induced by the possibility of large-sized variations in m (bank runs) suggests that this model has the potential to explain the evidence about the rise of excess reserve ratios in the US that I documented in the first part of this work. An evident limitation of the model is that the size of jumps is still very small compared to the observed variation in the data. Further investigation of the model could yield deeper insights about its theoretical properties, potentially leading to more appropriate parametrizations that produce larger jumps; at the same time, more accurate bank-level data on excess reserves, withdrawals and deposits could help to assess its appropriateness for the study of excess reserve management. Exploration of these issues is left for future research.

References

(1) Franklin Allen and Elena Carletti. "An overview of the crisis: Causes, consequences, and solutions". In: *International Review of Finance* 10.1 (2010), pp. 1–26 (cit. on p. 9).

(2) Fernando Alvarez and Francesco Lippi. "The demand of liquid assets with uncertain lumpy expenditures". In: *Journal of Monetary Economics* 60.7 (2013), pp. 753–770 (cit. on pp. 3, 4, 17, 23, 28, 33, 37, 40).

(3) Avner Bar-ilan, David Perry, and Wolfgang Stadje. "A generalized impulse control model of cash management". In: *Journal of Economic Dynamics and Control* 28.6 (2004), pp. 1013–1033 (cit. on p. 3).

(4) William Baumol. "The Transactions Demand for Cash: An Inventory Theoretic Approach". In: *The Quarterly Journal of Economics* 66.4 (1952), pp. 545–556 (cit. on p. 2).

(5) Jose M. Berrospide. "Bank liquidity hoarding and the financial crisis: an empirical evaluation". In: *Finance and Economics Discussion Series* 2013-03 (2013) (cit. on pp. 2, 14, 15).

(6) Ben R. Craig and Matthew Koepke. "Excess Reserves: Oceans of Cash". In: *Economic Commentary* Feb (2015), p. 2 (cit. on pp. 2, 11, 13).

(7) Alex Cukierman. "The Behavior of the Money Multiplier During and after the Subprime Crisis: Implications for the Transmission Mechanism of Monetary Policy". In: *CEPR Discussion Papers* Feb (2017) (cit. on p. 5).

(8) Avinash Dixit. "The Art of Smooth Pasting". In: *Fundamentals of Pure and Applied Economics*. Harwood Academic Publishers, 1993 (cit. on pp. 21, 33).

(9) Huberto Ennis and Alexander Wolman. "Large excess reserves in the U.S.: a view from the cross-section of banks". In: *Richmond FED Working Papers* 12-05 (2012) (cit. on pp. 2, 13, 16).

(10) Peter A. Frost. "Banks' Demand for Excess Reserves". In: *Journal of Political Economy* 79.4 (1971), pp. 805–825 (cit. on pp. 3, 14).

(11) Ricardo Lagos. *Inside and outside money*. Staff Report 374. Federal Reserve Bank of Minneapolis, 2006 (cit. on p. 32).

(12) Merton Miller and Daniel Orr. "A Model of the Demand for Money by Firms". In: *The Quarterly Journal of Economics* 80.3 (1966), pp. 413–435 (cit. on pp. 3, 17, 28).

(13) Frederic Minshkin and Eugene White. “Unprecedented Actions: The Federal Reserve’s Response to the Global Financial Crisis in Historical Perspective”. In: *NBER Working Paper Series* (2014) (cit. on pp. 9, 10).

(14) Frederic Mishkin. *The Economics of Money, Banking and Financial Markets*. Pearson Education Limited, 2016 (cit. on p. 5).

(15) Alberto M. Ramos. *Deposit insurance, bank capital structures and the demand for liquidity*. Working Paper Series, Issues in Financial Regulation WP-96-8. Federal Reserve Bank of Chicago, 1996 (cit. on pp. 13, 15).

(16) Rajdeep Sengupta and Yu-Man Tam. “The LIBOR-OIS spread as a summary indicator”. In: *Monetary Trends* Nov (2008) (cit. on p. 15).

(17) Nancy L Stokey. *The Economics of Inaction: Stochastic Control models with fixed costs*. Princeton University Press, 2008 (cit. on p. 21).

(18) Nancy L. Stokey, Edward C. Prescott, and Robert E. Lucas. *Recursive methods in economic dynamics*. 1989, xviii, 588 p. : (cit. on p. 26).

(19) James Tobin. “The Interest-Elasticity of Transactions Demand For Cash”. In: *The Review of Economics and Statistics* 38.3 (1956), pp. 241–247 (cit. on p. 3).

(20) John Walter and Renee Courtois. “The effect of interest on reserves on monetary policy”. In: *Richmond Fed Economic Brief* Dec (2009), pp. 09–12 (cit. on pp. 15, 32).

Sitography

(21) St. Louis FED. *FRED Economic Data*. URL: https://fred.stlouisfed.org/ (cit. on p. 32).

(22) Federal Reserve. *Policy Tools*. URL: https://www.federalreserve.gov/monetarypolicy/policytools.htm (cit. on p. 32).

(23) Federal Reserve. *What is the Money Supply? Is it Important?* URL: https://www.federalreserve.gov/faqs/money_12845.htm (cit. on p. 32).

A Appendix A

A.1 Data

All the data used in the empirical part is taken from FRED Economic Data (21).

A.2 Monetary aggregates and key interest rates in the US

In financially developed economies, the money supply is measured using a spectrum of measures which span all kind of assets that are classifiable as money, that is, that serve as a medium of exchange (Lagos (2006)). Such measures are called *monetary aggregates* and range from narrower ones, which include only the most liquid kinds of money (cash and assets that are easily and quickly convertible into cash) to broader ones, which include also less liquid assets.

Information about how the money supply is measured in the US is provided in the FED website (23), which defines the three measures of money cited above in the following way.

- The *monetary base* is defined as the sum of currency in circulation and reserve balances (deposits held by banks and other depository institutions in their accounts at the Federal Reserve).

- *M1* is defined as the sum of currency held by the public and transaction deposits at depository institutions.

In the United States there are three *key interest rates* that deserve to be mentioned for their impact on interbank lending and reserve-holding behaviour. These three rates are policy tools that the FED uses in order to implement its desired monetary policy stance (22).

- The *federal funds target range* is a range of values specified via a lower and upper limit established by the Federal Open Market Committee (FOMC), which sets the short-term objective for monetary policy.[8] As explained in Walter and Courtois (2009), open-market operations (sales and purchases of securities) are then used by the FED to adjust the supply of reserve balances so that it intersects bank's demand for reserves at a level at which the market clearing fed funds rate (measured by the *effective federal funds rate*, a weighted average of the interest rates at which depository institutions lend reserve balances to each other in the interbank market) stays within the target range.

- The *discount rate* is the interest rate charged to depository institutions on loans received from the FED's lending facility (the *discount window*). There are three discount windows, each one with its own discount rate: primary credit, secondary credit, seasonal credit. I will always refer to the discount rate that applies to the *primary credit program*, in which financially sound institutions borrow from the FED funds for a very short term. The discount rate is typically higher than the *effective fed funds rate*.

- The *interest rate on required and excess reserve balances* is the interest paid to depository institutions on balances they hold at Federal Reserve Banks. In principle, the rate on required and excess reserves can differ, but from December 2008 they have been equal.

[8] Until January 2009, the FOMC used to specify a federal funds target *rate*.

B Appendix B

B.1 Proof of Proposition 1

Observe that for $m \in (0, m^*)$, $T(m)$ can be described by the approximation

$$T(m) = \Delta + \frac{1}{2}T\left(m + \sqrt{\Delta}\sigma\right) + \frac{1}{2}T\left(m - \sqrt{\Delta}\sigma\right), \tag{15}$$

since in an interval of length Δ with equal probability the excess reserve holdings can increase or decrease by $\sqrt{\Delta}\sigma$. Expanding the RHS,

$$\begin{aligned} T(m) &\simeq \Delta + \frac{1}{2}\left(T(m) + T'(m)\left(\sqrt{\Delta}\sigma\right) + \frac{1}{2}T''\left(\sqrt{\Delta}\sigma\right)^2\right) + \\ &\quad + \frac{1}{2}\left(T(m) - T'(m)\left(\sqrt{\Delta}\sigma\right) + \frac{1}{2}T''\left(\sqrt{\Delta}\sigma\right)^2\right) = \\ &= \Delta + T(m) + \frac{1}{2}T''\left(\sqrt{\Delta}\sigma\right)^2, \end{aligned}$$

where the approximation holds exactly as $\Delta \to 0$. Simplifying yields

$$0 = \Delta + \frac{1}{2}T''(m)\Delta\sigma^2 \quad \Longrightarrow \quad 0 = 1 + \frac{\sigma^2}{2}T''(m).$$

Thus, we have a second order non-homogeneous differential equation, whose solution is $T(m)$. To have a unique solution, we need two boundary conditions, which are given by $T(0) = T(\bar{m}) = 0$. Hence, we have the Cauchy problem

$$\begin{cases} T''(m) = -\dfrac{2}{\sigma^2}, \\ T(0) = T(\bar{m}) = 0, \end{cases}$$

with solution

$$T(m) = \frac{m(\bar{m} - m)}{\sigma^2}.$$

The controlled process $m(t)$ is called a *regenerative process*, which means that there exist exist stopping times $0 \le \tau_0 < \tau_1 < \ldots$ such that the post-τ_n process $\{m(\tau_n + t)\,|t \ge 0\}$ is distributed exactly as the post-τ_0 process $\{m(\tau_0 + t)\,|t \ge 0\}$ and it is independent of previous values of the process. Each time interval between successive stopping times is called a *cycle* and each cycle starts at $m = m^*$. Hence, we can say that the expected length of a cycle is given by

$$T(m^*, \bar{m}) = \frac{m^*(\bar{m} - m^*)}{\sigma^2}, \tag{16}$$

the expected time before an adjustment when excess reserves are at the post-adjustment value. From (16), we can derive the expected number of adjustments per unit of time by

$$N(m^*, \bar{m}) = \frac{1}{T(m^*)} = \frac{\sigma^2}{m^*(\bar{m} - m^*)},$$

which is the desired result.

B.2 Proof of Proposition 2

Following Dixit (1993) and Alvarez and Lippi (2013) we can derive this stationary distribution using the *Kolmogorov forward equation*. To derive it, we can use the following discrete-time, discrete-state approximation for the law of motion of the density $h(m)$ with respect to m, taking the limit for $\Delta \to 0$.

$$h(m) = \frac{1}{2}h\left(m + \sqrt{\Delta}\sigma\right) + \frac{1}{2}h\left(m - \sqrt{\Delta}\sigma\right), \quad \text{for } m \in (0, m^*) \cup (m^*, \bar{m}) \tag{17a}$$

$$h(m^*) = \frac{1}{2}h\left(m^* + \sqrt{\Delta}\sigma\right) + \frac{1}{2}h\left(m^* - \sqrt{\Delta}\sigma\right) + \frac{1}{2}h\left(0\right) + \frac{1}{2}h\left(\bar{m}\right), \tag{17b}$$

$$h(0) = \frac{1}{2}h\left(\sqrt{\Delta}\sigma\right), \tag{17c}$$

$$h\left(\bar{m}\right) = \frac{1}{2}h\left(\bar{m} - \sqrt{\Delta}\sigma\right). \tag{17d}$$

Equation 17a states that the probability that the excess reserves are at a generic point m in the interior (with the exception of the target level m^*) is given by the sum of the probabilities that it was slightly above (slightly below) m a fraction Δ of time before, times the probability of a downward (upward) adjustment, which is $1/2$ (which follows from the fact that $\mu = 0$ in the unregulated process for excess reserves). The other conditions have a similar interpretation. A second order Taylor approximation of (17a) yields

$$\begin{aligned} h(m) \;&\simeq\; \frac{1}{2}\left(h\left(m\right) + h'(m)\left(\sqrt{\Delta}\sigma\right) + \frac{1}{2}h''(m)\left(\sqrt{\Delta}\sigma\right)^2\right) + \\ &+\frac{1}{2}\left(h\left(m\right) + h'(m)\left(-\sqrt{\Delta}\sigma\right) + \frac{1}{2}h''(m)\left(\sqrt{\Delta}\sigma\right)^2\right) = \\ &= h(m) + \frac{h''(m)}{2}\Delta\sigma^2. \end{aligned}$$

Dividing by Δ and simplifying yields

$$h''(m) = 0, \quad \text{for } m \in (0, m^*) \cup (m^*, \bar{m}).$$

Thus, the density is linear in the interior. A second order approximation of (17c) yields

$$\begin{aligned} h(0) \;&\simeq \frac{1}{2}\left(h\left(0\right) + h'(0)\left(\sqrt{\Delta}\sigma\right) + \frac{1}{2}h''(0)\left(\sqrt{\Delta}\sigma\right)^2\right) = \\ &= \frac{h(0)}{2} + \frac{h'(0)}{2}\sqrt{\Delta}\sigma + \frac{h''(m)}{2}\Delta\sigma^2. \end{aligned}$$

Taking the limit as $\Delta \to 0$, we get

$$h(0) = 0.$$

Repeating the same steps for $h(\bar{m})$, we easily get

$$h(\bar{m}) = 0.$$

To find h, we should solve the system

$$\begin{cases} h''(m) = 0, \quad \text{for } m \in (0, m^*) \cup (m^*, \bar{m}), \\ h(0) = 0, \\ h(\bar{m}) = 0, \\ \displaystyle\int_0^{\bar{m}} h(m)dm = 1, \end{cases} \tag{18}$$

where the last condition follows from the definition of probability density function. As pointed above, since $h''(m) = 0$ in the interior, h is a linear function in $(0, m^*) \cup (m^*, \bar{m})$. Since it should integrate to 1 and it cannot change slope in $(0, m^*) \cup (m^*, \bar{m})$, the only possibility is that it is linearly increasing in $(0, m^*)$ and linearly decreasing in $(m^*, \bar{m})$. Thereby, we write

$$h'(m) = \alpha, \quad h(m) = \alpha m + \beta, \quad \text{for } m \in (0, m^*),$$

for some $\alpha > 0, \beta \in \mathbb{R}$, and

$$h'(m) = \eta, \quad h(m) = \eta m + \gamma, \quad \text{for } m \in (m^*, \bar{m}),$$

for some $\eta < 0, \gamma \in \mathbb{R}$. Exploiting all the condition in (18) and imposing continuity at $0, m^*$ and $\bar{m}$ we obtain the solution

$$\begin{cases} h(0) = h(\bar{m}) = 0 \\ h(m) = \dfrac{2m}{m^*\bar{m}}, & \text{for} \quad m \in (0, m^*], \\ h(m) = \dfrac{2}{(\bar{m} - m^*)} - \dfrac{2m}{\bar{m}\,(\bar{m} - m^*)}, & \text{for} \quad m \in (m^*, \bar{m}), \end{cases}$$

which is a *triangular* probability distribution with lower limit 0, mode m^* and upper limit $\bar{m}$, as desired. Then, I write

$$M(m^*, \bar{m}) = \mathbb{E}\,(m) = \int_0^{\bar{m}} m h(m) dm.$$

The expectation may be decomposed as follows

$$\int_0^{\bar{m}} m h(m) dm = \int_0^{m^*} m \cdot \left(\frac{2m}{m^*\bar{m}}\right) dm + \int_{m^*}^{\bar{m}} m \cdot \left(\frac{2}{(\bar{m} - m^*)} - \frac{2m}{\bar{m}\,(\bar{m} - m^*)}\right) dm.$$

Straightforward integration yields

$$M(m^*, \bar{m}) = \frac{\bar{m} + m^*}{3},$$

as desired.

B.3 Proof of Proposition 4

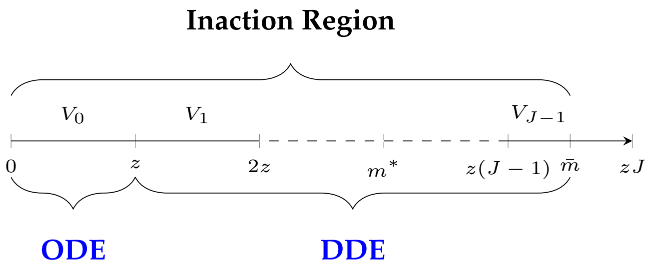
From the ODE (11a) and the DDE (11b), it is possible to recover an equivalent system of J ODEs. Taking m^* and $\bar{m}$ as given, it is possible to split the interval $[0, \bar{m}]$ into J sub-intervals. The first $J - 1$ intervals are of size z and given by

$$[zj, z(j+1)]\,, \quad \text{for } j = 0, 1, ..., J-2,$$

while the Jth interval is given by

$$[z(J-1), \min\{\bar{m}, zJ\}]\,.$$

I also denote by $j^* \in \mathbb{N}$ the smallest integer such that $z(j^*+1) \geq m^*$. Hence, the (j^*+1)th interval contains m^*, the optimal return point. I index the solution of the ODE in each interval by j. In the first interval $[0, z]$, the solution is described by (11a). Hence, denoting



Graphical representation of the solution method.

its solution by $V_0 : [0, z] \to \mathbb{R}$, I can write

$$(r+\kappa)V_0(m) = Rm + \frac{V_0''(m)\sigma^2}{2} + \kappa(V^* + b), \tag{19}$$

for $0 \le m \le z$. For $j = 1, ..., J$, taking as given the function $V_{j-1}(\cdot)$, I have the following ODE for $V_j : [zj, \min\{z(j+1), \bar{m}\}] \to \mathbb{R}$,

$$(r+\kappa)V_j(m) = Rm + \frac{V_j''(m)\sigma^2}{2} + \kappa V_{j-1}(m-z). \tag{20}$$

The system composed by the ODE (19) and by the $J-1$ ODEs (20), has associated boundary conditions

$$V_0(0) = V^* + b, \tag{21a}$$
$$V_{J-1}(\bar{m}) = V^* + b, \tag{21b}$$
$$V_{j^*}(m^*) = V^*, \tag{21c}$$
$$V_{j-1}(zj) = V_j(zj), \quad \text{for } j = 1, 2, ..., J-1, \tag{21d}$$
$$V_{j-1}'(zj) = V_j'(zj), \quad \text{for } j = 1, 2, ..., J-1, \tag{21e}$$

where (21a), (21b) and (21c) derive respectively from Equation 13a, Equation 13b and Equation 10. The solution to is given by

$$V(m, m^*, \bar{m}) = V_j(m), \quad \text{for } m \in [zj, \min\{z(j+1), \bar{m}\}].$$

The solution method is to solve for V_0, then use the solution to solve for V_1 and so on. In solving for V_0, we will have three unknown parameters. Using boundary conditions, we find $V_0, V_1, ..., V_{J-1}$ as functions of these three parameters. Using the three boundary conditions in Equation 21 we can pin down these three parameters. After that, having completely characterized the value function $V(m, m^*, \bar{m})$, one can use respectively Equation 12 and Equation 14 to obtain the optimal target and trigger values. Start from Equation 19. The homogeneous solution is

$$V_0(m)^h = B_{0,0}^1 e^{\lambda_1 m} + B_{0,0}^2 e^{\lambda_2 m},$$

where $\lambda_1 > 0 > \lambda_2$ are the two solutions to $r + \kappa = (\sigma^2/2)\lambda^2$. The particular solution is given by

$$V_0(m)^p = A_0 + D_0 m.$$

Hence, the solution is

$$V_0(m) = A_0 + D_0 m + B_{0,0}^1 e^{\lambda_1 m} + B_{0,0}^2 e^{\lambda_2 m}.$$

Substituting into Equation 19 we get

$$(r+\kappa)\left(A_0 + D_0 m\right) = \kappa(V^* + b) + Rm,$$

because the exponential terms vanish. Matching coefficients we immediately get

$$D_0 = \frac{R}{r+\kappa}, \tag{22}$$

$$A_0 = \frac{\kappa(V^*+b)}{r+\kappa}. \tag{23}$$

From Equation 21a we get

$$A_0 + \sum_{k=1}^{2} B^k_{0,0} = V^* + b. \tag{24}$$

Combining Equation 23 and Equation 24 we can express A_0 and V^* as functions of known parameters and of the coefficients B^k_{00}, getting

$$A_0 = \frac{\kappa(B^1_{0,0} + B^2_{0,0})}{r}, \quad \text{and} \quad V^* = \frac{(r+\kappa)(B^1_{0,0} + B^2_{0,0})}{r} - b.$$

Now, following Alvarez and Lippi (2013), I guess a solution for V_j of the form

$$V_j(m) = A_j + D_j(m - zj) + \sum_{k=1}^{2}\sum_{i=0}^{j} B^k_{j,i} e^{\lambda_k(m-zj)} \cdot (m-zj)^i. \tag{25}$$

In the next step, I will check that this guess works and I will pin down the coefficients $\{A_j\}_{j=1}^{J-1}$, $\{D_j\}_{j=1}^{J_1}$ and $\{B^k_{j,i}\}_{j=1}^{J-1}$ as functions of $B^1_{0,0}$ and $B^2_{0,0}$, exploiting the recursive nature of the problem and known initial conditions A_0 and D_0.

Plugging the guessed solution in Equation 25 into the ODEs in (20), we obtain

$$\begin{aligned}
&(r+\kappa)\left(A_j + D_j(m-zj) + \sum_{k=1}^{2}\sum_{i=0}^{j} B^k_{j,i} e^{\lambda_k(m-zj)} \cdot (m-zj)^i\right) \\
= \;& Rm + \kappa\left(A_{j-1} + D_{j-1}(m-zj) + \sum_{k=1,2}\sum_{i=0}^{j-1} B^k_{j-1,i} e^{\lambda_k(m-zj)} (m-zj)^i\right) + \\
+ \;& \frac{\sigma^2}{2}\left(\sum_{k=1,2}\sum_{i=0}^{j} B^k_{j,i} e^{\lambda_k(m-zj)}\left(\lambda_k^2 (m-zj)^i + 2\lambda_k i (m-zj)^{i-1} + i(i-1)(m-zj)^{i-2}\right)\right)
\end{aligned} \tag{26}$$

Matching the coefficients for m in (26) I obtain a recursive expression for D_j

$$D_j = \frac{R}{r+\kappa} + \frac{\kappa}{r+\kappa} D_{j-1}, \quad \text{for } j = 1, 2, ..., J-1. \tag{27}$$

Hence, starting from the known value D_0 I can pin down all $\{D_j\}_{j=0}^{J-1}$. Matching the constant in the same equation I obtain a recursive expression for A_j.

$$A_j = \frac{\kappa A_{j-1} - \kappa z j D_{j-1}}{r+\kappa} + D_j z j, \quad \text{for } j = 1, 2, ..., J-1. \tag{28}$$

Matching the coefficients for $e^{\lambda_k(m-zj)}(m-zj)^j$ we don't get any restriction. Matching the coefficients for $e^{\lambda_k(m-zj)}(m-zj)^{j-1}$, we get a difference equation for $B^k_{j,j}$, i.e.,

$$B^k_{j,j} = -\frac{\kappa}{\sigma^2\lambda_k} B^k_{j-1,j-1}, \quad \text{for } j = 1, 2, ..., J-1, k = 1, 2. \tag{29}$$

Using Equation 29 we can solve for $\{B^k_{j,j}\}_{j=0}^{J-1}$ given $B^k_{0,0}$ for $k = 1, 2$. Likewise the coefficients for $e^{\lambda_k(m-zj)}(m-zj)^i$, for $i = 0, 1, ..., j-2$ we get

$$-\sigma^2\lambda_k(i+1)B^k_{j,i+1} = \kappa B^k_{j-1,i} + \frac{\sigma^2}{2} B^k_{j,i+2}(i+2)(i+1), \quad \text{for } j = 2, ..., J-1, k = 1, 2. \tag{30}$$

I also have Equation 21d and Equation 21e to exploit. These conditions imply

$$A_j + \sum_{k=1}^{2} B^k_{j,0} = A_{j-1} + D_{j-1}z + \sum_{k=1}^{2}\sum_{i=0}^{j-1} B_{j-1,i}e^{\lambda_k z}z^i, \tag{31a}$$

$$D_j + \sum_{k=1}^{2} B^k_{j,0}\lambda_k + \sum_{k=1}^{2} B^k_{j,1} = D_{j-1} + \sum_{k=1}^{2}\sum_{i=0}^{j-1} B^k_{j-1,i}e^{\lambda_k z}\left[\lambda_k z^i + iz^{i-1}\right]. \tag{31b}$$

With all the equations I have until now, I can solve for $\{A_j, D_j, B^k_{j,i}\}$ as functions of $B^k_{0,0}$. To pin down the values $B^k_{0,0}$ I can exploit Equation 21b and Equation 21c.

$$V^* = A_{j^*} + D_{j^*}(m^* - zj^*) + \sum_{k=1}^{2}\sum_{i=0}^{j^*} B^k_{j^*,i}e^{\lambda_k(m^*-zj^*)}(m^*-zj^*)^i, \tag{32a}$$

$$V^* + b = A_{J-1} + D_{J-1}(\bar{m} - z(J-1)) + \sum_{k=1}^{2}\sum_{i=0}^{J-1} B^k_{J-1,i}e^{\lambda_k(\bar{m}-z(J-1))}(\bar{m} - z(J-1))^i. \tag{32b}$$

Having solved also for $B^k_{0,0}$, the function $V(m, m^*, \bar{m})$ is now completely characterized.

B.4 Proof of Proposition 6

As $\rho \to 0$ (i.e., as the future periods become all equally weighted, which is what I need since I look for historical averages), the initial condition $m(0) = m$ doesn't matter anymore, so that aggregate statistics are given by

$$M = \lim_{\rho\to 0} M(m), \quad \text{and} \quad n = \lim_{\rho\to 0} n(m),$$

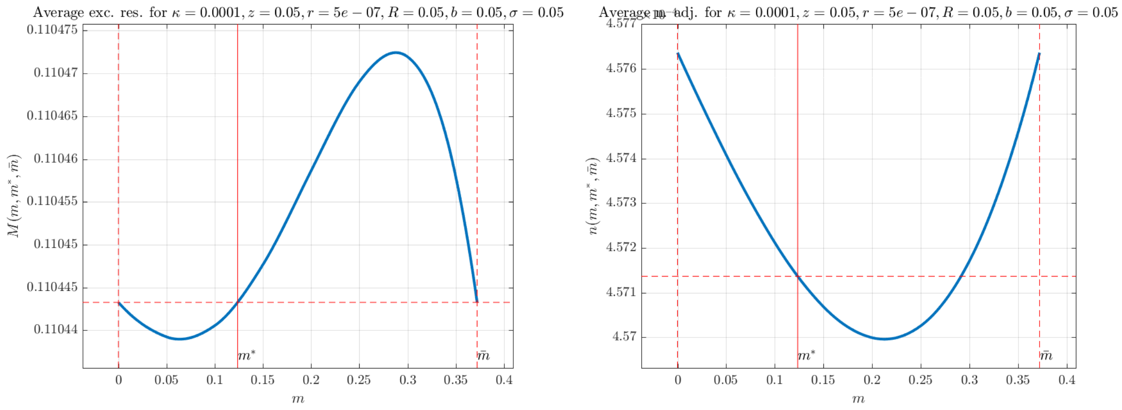
To retrieve $n(m)$ and $M(m)$, I can use a discrete-time discrete-state approximation to derive the HJB equations for these functions. More specifically, start respectively from

$$M(m) = \rho m\Delta + \frac{1}{1+\rho\Delta}\left(\kappa\Delta M(m-z) + (1-\kappa\Delta)\left(\frac{1}{2}M(m+\sqrt{\Delta}\sigma) + \frac{1}{2}M(m-\sqrt{\Delta}\sigma)\right)\right)$$

for $M(m)$ and from

$$n(m) = \rho\kappa\Delta + \frac{1}{1+\rho\Delta}\left(\kappa\Delta n(m-z) + (1-\kappa\Delta)\left(\frac{1}{2}n(m+\sqrt{\Delta}\sigma) + \frac{1}{2}n(m-\sqrt{\Delta}\sigma)\right)\right)$$

for $n(m)$. Then, use Taylor expansions and Ito's Lemma, divide by Δ and let $\Delta \to 0$ to obtain the following ODE-DDE systems for $M(m)$ and $n(m)$, with associated boundary



(a) Average excess reserve ratio $M(m, m^*, \bar{m})$ where $(m^*, \bar{m})$ is the optimal policy for displayed parameter values.

(b) Average number of adjustments $n(m, m^*, \bar{m})$ where $(m^*, \bar{m})$ is the optimal policy for displayed parameter values.

conditions.

$$\begin{cases} \rho M(m) = \rho m + \dfrac{M''(m)\sigma^2}{2} + \kappa\left(M(m^*) - M(m)\right) & \text{for } 0 < m \leq z < \bar{m}, \\ \rho M(m) = \rho m + \dfrac{M''(m)\sigma^2}{2} + \kappa\left(M(m-z) - M(m)\right), & \text{for } 0 < z < m < \bar{m}, \\ M(0) = M(\bar{m}) = M(m^*). \end{cases}$$

$$\begin{cases} \rho n(m) = \rho\kappa + \dfrac{n''(m)\sigma^2}{2} + \kappa\left(n(m^*) - n(m)\right) & \text{for } 0 < m \leq z < \bar{m}, \\ \rho n(m) = \rho\kappa + \dfrac{n''(m)\sigma^2}{2} + \kappa\left(n(m-z) - n(m)\right), & \text{for } 0 < z < m < \bar{m}, \\ n(0) = n(\bar{m}) = \rho + n(m^*), \end{cases}$$

Note that this ODE-DDE systems are similar to the system (11a)-(11b), with few differences. Indeed, the way of solving them is identical and the solution for $n(m)$ and $M(m)$ is the same described in Proposition 4, except for the coefficients that will be obviously different. To get a close approximation of n and M, I solve numerically these systems for $\rho \simeq 0$ after having computed the optimal policy $(m^*, \bar{m})$, and I obtain $N = N(m^*) \simeq N(m), \forall m$ and $M = M(m^*) \simeq M(m), \forall m$. In Figure 13a and Figure 13b I display the functions $M(m, m^*, \bar{m})$ and $n(m, m^*, \bar{m})$.

B.5 Large-Sized Jumps and Piecewise-Defined Policy Rules

Consider a commercial bank whose excess reserve ratio decreases by $\eta = \mu + kz$ per unit of time. The exogenous process for the excess reserve ratio has two components, one continuous at rate μ and the other discontinuous, with jumps of size z each $1/\kappa$ periods of time. The commercial bank must decide how frequently to adjust upwards the excess reserve ratio (n) and by how much (U), to solve

$$\min_{M,N} RM + bn, \quad \text{s.t.} \quad nU = \mu + \kappa z, \tag{33}$$

where U is the size of the upward adjustment. As the continuous component μ is not needed to obtain piecewise-defined policies, I focus on the case $\mu \to 0$. The optimal policy is described in the following Proposition.

Proposition 7 *Let $\mu \to 0$. Then, the number of adjustments per unit of time n, the average excess reserve ratio M and the average size of upward adjustments U are given by*

$$\text{If } \kappa < \frac{Rz}{2b} \quad \Rightarrow \quad n = \kappa, \quad M = 0, \quad U = z.$$
$$\text{If } \kappa \geq \frac{Rz}{2b} \quad \Rightarrow \quad n = \sqrt{\frac{Rz\kappa}{2b}} \leq \kappa, \quad M = \frac{z}{2}\left(\frac{\kappa}{n} - 1\right) \geq 0, \quad U = \sqrt{\frac{2b\kappa z}{R}} \geq z.$$

Proof. A solution of the model with $\mu \neq 0$ is available in Alvarez and Lippi (2013). Then, taking the limits for $\mu \to 0$ the result is easily obtained. ■

When κ is small, adjustments coincide exactly with jumps ($n = \kappa$) and M is equal to zero. When κ is large, meaning that jumps are relatively frequent, it will be optimal to have less adjustments than jumps per unit of time ($n < \kappa$) and the average excess reserve ratio M will be higher than zero. The key idea that this simplified deterministic model conveys is that even small changes in κ may trigger shifts in the optimal policy followed by a commercial bank. In other words, for certain changes in κ (across the threshold $Rz/2b$)[9], the elasticity of M to κ is degenerate, as M does not vary smoothly with κ but it suddenly changes as the commercial bank modify its policy. This feature is a consequence of the presence of large-sized jumps: in a deterministic model without jumps (a simple Baumol-Tobin model) piecewise-defined policy rules do not arise.

C Appendix C

In this last Appendix I describe the Matlab routine I implemented to obtain the results, which can perform mainly two tasks.

1. For given parameters of the model, it finds the optimal policy $(m^*, \bar{m})$, it checks if it corresponds to a global minimum and it computes the value function $V(m, m^*, \bar{m})$. I describe the code that solves the problem in Appendix C.1.

2. It finds the average number of adjustments per unit of time $N(m^*, \bar{m})$ and the average excess reserve ratio $M(m^*, \bar{m})$ and it simulates a stochastic process subject to the optimal control policy in order to obtain an estimate of $h(m)$, the invariant distribution of excess reserve ratios. I describe the code that performs these tasks in Appendix C.2.

C.1 Finding Optimal Policies

To find the optimal policy, I exploited two functions, *fun_findV* and *fun_Vstar*. Given a certain policy vector $(m^*, \bar{m})$, these two functions combined yield the function $V(m, m^*, \bar{m})$ described in 4 and its minimum $V(m^*, m^*, \bar{m})$. Then, a standard built-in optimization routine (*fmincon*) is exploited to find the optimal policy vector $(m^*, \bar{m})$.

More specifically, the function *fun_findV* takes three vectors as inputs: the policy vector **policy** $= (m^*, \bar{m})$, the vector of initial guesses for V^* and $B^1_{0,0}$, **input** $= (V^*_{init}, B^1_{0,0\ init})$ and the parameter vector **param**.

[9] It is important to notice that there is substitutability between κ and z. Indeed, the optimal policy switches when $\kappa/z = R/2b$. Hence, for a given jump size z, if the jump frequency κ increases it is more likely that the bank follows a policy with high M (*front loading*). On the contrary, for a given frequency κ, if the jump size z increases then it is more likely that the bank follows a policy with low M (*no front loading*).

```
%I define the function and all of its inputs.
function[zero,B001,B002,J,Jstar,A0,D0,A,D,B01,B02,B1,B2]=fun_findV(
    input,policy,param)
global kappa z r sigma lambda1 lambda2
alpha_0 = param(1);
nu_0 = param(2);
alpha    = param(3);
nu   = param(4);
alphastar= param(5);
Vstar=input(1);
B001=input(2);
mstar=policy(1);
mbar=policy(2);
```

The parameter vector will change according to what we want to compute among V, M or n. To compute V, I will input *param_V* and so on.

```
param_V=[((kappa*b)/r)    R/r     0    R/r     b/r];
    % Parameters given to fun_findV to find V
param_M=[0                1       0    1       0];
    % Parameters given to fun_findV to find M
param_n=[kappa            0       0    0       1];
    % Parameters given to fun_findV to find N
```

Given the values for m^* and $\bar{m}$, it is immediately possible to find J and j^*. Moreover, given the two initial guesses, I can compute immediately $B^2_{0,0}$, A_0 and D_0 using Equation 23, Equation 22 and Equation 24. Then, I create empty matrices for all coefficients of the value function, which I will fill with subsequent loops.

```
%I derive the number of intervals, based on values for mbar/mstar.
J=round(mbar/z);
if J<mbar/z || J==0
    J=J+1;
end
Jstar=round(mstar/z);
if Jstar>mstar/z
    Jstar=Jstar-1;
end
%Formulas for j=0;
D0=(r/(r+kappa))*nu_0;
A0=((kappa*Vstar)/(r+kappa))+alpha_0*(r/(r+kappa));
B002=Vstar+(r*alphastar)-A0-B001;
%Other coefficients of the solution, stored in matrices and vectors
B1=zeros(J-1,J-1);
B01=zeros(J-1,1);
B2=zeros(J-1,J-1);
B02=zeros(J-1,1);
A=zeros(J-1,1);
D=zeros(J-1,1);
```

First, using Equation 27, Equation 28 and Equation 29 I can compute $\{A_j\}_{j=1,\dots,J-1}$, $\{D_j\}_{j=1,\dots,J-1}$ and the diagonal of $\mathbf{B_k}$, i.e., I can find $\{B^k_{j,j}\}_{j=1,\dots,J-1}$ for $k = 1, 2$. Then, to find all coefficients in $\mathbf{B_{0k}}$ and $\mathbf{B_k}$, I exploit the following procedure. First, observe that I have $B^k_{0,0}$ and all the diagonal $\{B^k_{j,j}\}_{j=1,\dots,J-1}$ for $k = 1, 2$. Note that Equation 31a and Equation 31b form a two-dimensional system of linear equations. If I have $\{B^k_{j-1,i}\}_{i=0,1,\dots,j-1}$, I can recover $B^k_{j,0}$. At any iteration of the loop, since I have all the required coefficients, I can compute $B^k_{j,0}$. Having $B^k_{j,0}$, $B^k_{j,j}$ and all the $\{B^k_{j-1,i}\}_{i=0,1,\dots,j-1}$, I can exploit Equation 30 to determine all coefficients $\{B^k_{j,i}\}_{i=1,...,j-1}$. In this way, I obtain all the coefficients needed. The following loop exploits the aforementioned algorithm.

```
%Here I compute {A_j}, {D_j} and the diagonal {B_{j,j}}
for j=0:J-2
    if j==0  % this if takes care of the case of j=0;
        D(j+1)=(R/(r+kappa))+(kappa/(r+kappa))*D0;
        A(j+1)=((kappa*A0-z*kappa*(j+1)*D0)/(r+kappa))+D(j+1)*z*(j+1);
        B1(j+1,1)=-(kappa/(lambda1*(sigma^2)))*B001;
        B2(j+1,1)=-(kappa/(lambda2*(sigma^2)))*B002;
    else
        D(j+1)=(R/(r+kappa))+(kappa/(r+kappa))*D(j);
        A(j+1)=((kappa*A(j)-z*kappa*(j+1)*D(j))/(r+kappa))+D(j+1)*z*(j
            +1);
        B1(j+1,j+1)=-(kappa/(lambda1*(sigma^2)*(j+1)))*B1(j,j);
        B2(j+1,j+1)=-(kappa/(lambda2*(sigma^2)*(j+1)))*B2(j,j);
    end
end
for j=0:J-2
    %Here I impose VM and SP to compute B_{1,0}.
    if j==0
        par={A0,A(1),D0,D(1),B001,B002,B1,B2};
        x=solvesys0(par);
        B01(1)=x(1);
        B02(1)=x(2);
    else
        %Having B_{j+1,j+1} as BCs, I recursively compute B_{j+1,i}
            until
        %i=2
        for i=j:-1:2
             B1(j+1,i)=-(kappa/(lambda1*(sigma^2)*i))*B1(j,i-1)-(1/(
                 lambda1*2))*B1(j+1,i+1)*(i+1);
             B2(j+1,i)=-(kappa/(lambda2*(sigma^2)*i))*B2(j,i-1)-(1/(
                 lambda2*2))*B2(j+1,i+1)*(i+1);
        end
        %Then, I compute B_{j+1,1}
        B1(j+1,1)=-(kappa/(lambda1*(sigma^2)))*B01(j)-(1/lambda1)*B1(j
            +1,2);
        B2(j+1,1)=-(kappa/(lambda2*(sigma^2)))*B02(j)-(1/lambda2)*B2(j
            +1,2);

        %To end up, I compute B_{j+1,0}
        par={A,D,B01,B02,B1,B2,j};
        x=solvesys(par);
        B01(j+1)=x(1);
        B02(j+1)=x(2);

    end

end
end
```

The inner functions *solvesys0* and *solvesys* solve the two-dimensional system of linear equations at each iteration of the loop, using *linsolve*. Now, I have all the coefficients I needed. Then, using Equation 32a and Equation 32b I obtain two implied values for V^*, $V^*_{imp_1}$ and $V^*_{imp_2}$. The code below compute these two values, *imp1_Vstar* and *imp2_Vstar*. Once that I have $V^*_{imp_1}$ and $V^*_{imp_2}$ I can compute the output of the function, the vector **zero** $= (V^*_{imp_1} - V^*_{init}, V^*_{imp_2} - V^*_{init})$ which measures the difference between implied values and guessed value for V^*.

```
%Now I have to compute the implied Vstars
if Jstar==0
    imp1_Vstar = A0 + D0*(mstar) + B001*exp(lambda1*mstar)+B002*exp(
        lambda2*mstar);
elseif Jstar>0
    imp1_Vstar = A(Jstar) + D(Jstar)*(mstar-z*(Jstar)) + B01(Jstar)*exp
        (lambda1*(mstar-z*(Jstar)))+...
            +B02(Jstar)*exp(lambda2*(mstar-z*(Jstar))) ;
        for i=1:Jstar
        imp1_Vstar= imp1_Vstar + B1(Jstar,i)*exp(lambda1*(mstar-z*(
            Jstar))) * (mstar-z*(Jstar))^i +...
                +B2(Jstar,i)*exp(lambda2*(mstar-z*(Jstar))) * (mstar-z
                    *(Jstar))^i;
        end
end

if J==1
    imp2_Vstar = A0+D0*(mbar)+B001*exp(lambda1*mbar)+B002*exp(lambda2*
        mbar)-b;
elseif J>1
    imp2_Vstar =  A(J-1) + D(J-1)*(mbar-z*(J-1)) + B01(J-1,1)*exp(
        lambda1*(mbar-z*(J-1)))+...
            +B02(J-1,1)*exp(lambda2*(mbar-z*(J-1)))-b;
        for i=1:J-1
        imp2_Vstar= imp2_Vstar + B1(J-1,i)*exp(lambda1*(mbar-z*(J-1)))
            * (mbar-z*(J-1))^i +...
                +B2(J-1,i)*exp(lambda2*(mbar-z*(J-1))) * (mbar-z*(J-1))
                    ^i;
        end
end
% Outputs of the function
zero(1)=(imp1_Vstar-Vstar);
zero(2)=(imp2_Vstar-Vstar);
```

The function *fun_Vstar* takes as input the policy vector **policy** $= (m^*, \bar{m})$ and the parameter vector *param_V*. It finds $(V^*_{init}, B^1_{0,0\ init})$ such that **zero** $= (V^*_{imp_1} - V^*_{init}, V^*_{imp_2} - V^*_{init}) \simeq (0, 0)$. The function outputs $V(m^*, m^*, \bar{m})$ (the minimum of the value function given by following a certain policy) and all the coefficients A_j, D_j and $B^k_{j,i}$. Given these coefficients, the function *fun_F* computes the value function.

Exploiting *fun_findV* and *fun_Vstar* I am able to obtain the minimum of the value function for any combination of $(m^*, \bar{m})$. Then, using the built-in minimization routine *fmincon* (with $\bar{m} > m^*$ and Equation 21e) as constraints) I can find the optimal policy $(m^*, \bar{m})$.

C.2 Computing Aggregate Statistics

To find the functions $M(m)$ and $n(m)$ it is sufficient to exploit the routine *fun_vstar* using the optimal policy $(m^*, \bar{m})$ as a policy input, with the parameter vector **param** adjusted to solve the ODE-DDEs for M and n instead of the one for V. The routine yields the coefficients that characterize the shape of the functions $M(m)$ and $n(m)$. To obtain M and n , it is sufficient to get $M(m)$ and $n(m)$ for a very small discount factor (sort of taking the limits for $\rho \to 0$).

To obtain an approximation of $h(m)$, I simulate the stochastic process subject to the optimal control policy previously found. To do so, I developed a simple routine called *simulate*, which takes as inputs the vector **policy** which yields a vector **path** of the desired size **T** that contains the simulated controlled stochastic process, as shown below.

```
% This functions simulates the process dm(t)=\sigma dW(t)-zdN(t)
```

```
    subject to (m^{\ast},\bar{m}).
function [path]=simulate(policy,T)
global sigma kappa z
mstar=policy(1);
mbar=policy(2);
D=1/365;
path=zeros(round(T/D),1);
path(1)=(mbar-mstar)*rand+mstar;
for t=1:round(T/D)
    a1=randn;
    a2=rand;
    if a2<kappa
        if path(t)>z
            path(t+1)=path(t)-z;
        else
            path(t+1)=mstar;
        end
    else
        if path(t)+sigma*a1*sqrt(D)>=mbar || path(t)+sigma*a1*sqrt(D)
            <=0
            path(t+1)=mstar;
        else
            path(t+1)=path(t)+sigma*a1*sqrt(D);
        end
    end
end
```

All the Matlab code I developed for this work, comprehensive of the functions *fun_findV*, *fun_Vstar*, *solvesys0*, *solvesys*, *fun_F*, *simulate* and of the two main scripts that respectively find the optimal policy and perform comparative statics are available online at this link.