Department of Economics

Economic Analysis Research Group (EARG)

University of Reading

Money Creation in Russia: Does the Money Multiplier Exist?

by Vadim O. Grishchenko, Alexander Mihailov and Vasily N. Tkachev

Discussion Paper No. 2021-15

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

Money Creation in Russia: Does the Money Multiplier Exist?

Vadim O. Grishchenko* Alexander Mihailov† Vasily N. Tkachev‡

July 2021§

Abstract

For decades, the monetary economics literature has considered multiple deposit expansion via the money multiplier logic as empirically corroborated. However, the developments witnessed in advanced economies since the Global Financial Crisis challenged this settled view, and central banks as well as the Bank for International Settlements were among the first to openly reconsider it. In this paper, we revisit the issue empirically, but in a way aligned with a 'narrative' context of the evolving institutional frameworks for banking activities and monetary policy that profoundly and ultimately shape it out. Using a vector autoregression model estimated on Russian monthly data over two subsamples, 2005-2012 and 2012-2019, we find robust evidence that, while multiple deposit expansion may have existed in underdeveloped financial systems in the past, where the volume of lending was limited by the supply of bank reserves, nowadays lending is constrained mainly by the demand for credit. The key explanations we propose are: the rapid rise of money markets in the 20th-21st centuries, the unlimited access to central bank liquidity provision facilities, and the evolution of bank management from the 'golden rule' of banking, where liquidity gaps aim at zero, to Asset and Liability Management, where banks flexibly manage liquidity gaps. Our results robustly show that the influence on real money balances of money supply factors, such as bank reserve requirements and the real monetary base, has become statistically insignificant over the recent decade in Russia, while that of money demand factors, such as the nominal interest rate, has remained significant and negative, which is consistent with the economic intuition we have suggested.

Keywords: multiple deposit expansion, money multiplier, supply of bank reserves, demand for credit, evolution of bank management, monetary policy, Russia

JEL Classification Codes: E41, E42, E44, E51, E58, G21

*Corresponding author; Research and Forecasting Department, Bank of Russia; Moscow, Russia; grishchenkovo@cbr.ru; gvo@mail.cbr.ru; grishchenkovadim@yandex.ru

†Department of Economics, University of Reading, Whiteknights, Reading, Berkshire, RG6 6AA, UK; a.mihailov@reading.ac.uk

‡International Finance Department, Moscow State University of International Relations (MGIMO-University), Moscow, Russia; tkachev_mgimo@mail.ru

§First draft: October 2019. We are grateful to A. Faudot, N. Nenovsky and J. Kregel for comments and suggestions on earlier versions. All remaining errors or misinterpretations are the sole responsibility of the authors, and in particular do not represent views of the Bank of Russia.

Contents

List of Tables

List of Figures

1 Introduction: The Money Multiplier Concept – Famous but Dubious

For decades, the monetary economics literature has considered multiple deposit expansion by (commercial) banks via the mechanism of the so-called 'money multiplier' as empirically corroborated. However, the developments witnessed in advanced economies since the Global Financial Crisis (GFC) challenged this settled view, and central banks as well as the Bank for International Settlements were among the first to openly reconsider it – see, e.g., Borio and Disyatat (2009), Disyatat (2010), ECB (2011), Jakab and Kumhof (2015), Jordan (2018), Borio (2019).

There exist several different, though closely related, definitions of the concept of the money multiplier. In a simplest statement, the money multiplier under a fractional-reserve banking system can be defined as just the reciprocal of the required reserves ratio, which is a particular percentage of the deposit base of commercial banks that needs to be held on reserve accounts with the central bank according to the laws and regulations in a given country. This definition is equivalent to expressing a measure of money – say, the monetary aggregate denoted as 'broad money', $M2$ – as the product of the 'monetary base'[1] (treated in an accounting context as the liabilities of a central bank), MB, and the (corresponding to that measure of money) multiplier, $m2$ in this case, and appears in standard textbooks – e.g., Mishkin (2014):

$$M2 = m2 \times MB$$

Then, one can rewrite the (respective) multiplier as a ratio of the (respective) measure of money (a monetary aggregate), in the numerator, and the monetary base in the denominator:

$$m2 = \frac{M2}{MB}$$

One important difference in the variants of the money multiplier concept that are common in the literature arises from what measure of money ('narrow money' $M1$ or $M2$ or other) is taken in the numerator of its defining ratio. That is, $m1$, the narrow money multiplier, differs from $m2$, the broad money multiplier, in as much as $M1$ differs from $M2$.[2] A second difference encountered in research on the money multiplier is that often it is given in levels, as in most textbooks and the above definition, but sometimes it is written instead as a ratio of absolute changes, i.e., first differences – e.g., Carpenter and Demiralp (2012), eq. (1), p. 59: they express the (broad) money multiplier ($m2$) as a ratio of (the change, Δ, in) $M2$ to (the change in) MB, so that:

$$m2 = \frac{\Delta M2}{\Delta MB}$$

The most popular interpretation of the money multiplier states that it shows how much money is created by the banking system of a country for every unit of base money created by its central bank. The value of the money multiplier differs considerably for different countries at a given time and for different time periods in the same country. For example, in the United States (US) the money multiplier of broad money M2 to the monetary base was 10.6 in 1930, 4.5 in 1940, and 12.1 in 1985. The latest data on the money multiplier for several major economies is provided in Table 1. The table shows that Australia is an 'outlier' in this context, with its money multiplier being 3 to 5 times higher than that of the other countries. This extreme is not surprising for monetary economists and central bankers, and the reason for it is institutional, not

[1] Also referred to as 'base money' or 'central bank money' or 'high-powered money'.

[2] Central banks provide explicit and precise definitions of what is included in each monetary aggregate, such as $M1$ or $M2$, and monetary statistics regarding its evolution, as well as the evolution of the monetary base MB and the corresponding money multipliers $m1$ or $m2$, is usually published monthly.

empirical: simply, there are no reserve requirements in Australia. As a result, banks hold just as much reserves on accounts at the Reserve Bank of Australia as they need for transactions. On the other extreme is Japan, where the money multiplier has declined from 4.4 in 2014 to 2.7 in 2019, and Russia comes close to Japan.

Selected major world economies	2014	2015	2016	2017	2018	2019
Australia	25.9	25.4	25.8	26.2	25.8	n/a
China, P.R.: Mainland	4.2	5.0	5.0	n/a	n/a	n/a
Euro Area	7.1	5.5	4.4	3.7	3.8	4.0
India	5.5	5.3	8.1	5.8	5.4	5.3
Japan	4.4	3.5	3.0	n/a	2.7	2.7
Russian Federation	3.8	4.7	4.3	3.7	3.8	3.8
United Kingdom	6.9	6.8	6.3	n/a	n/a	n/a
United States	3.9	4.5	4.9	4.6	5.4	5.6

Note: n/a = the data are not available. Source: authors calculations of simple $m2$ ratios (not Δ) based on the IMF International Financial Statistics Database: https://data.imf.org/regular.aspx?key=61545868

Table 1: Broad Money to Monetary Base Ratio for Selected Countries, 2014-2019

A further decomposition of the money supply into currency in circulation (CC) and deposits of the private sector with banks (BD), and of the monetary base into currency in circulation and bank reserves (BR), yields the well-known alternative equation for the key components of the money multiplier:

$$m2 = \frac{CC + BD}{CC + BR} = \frac{1}{\frac{CC}{M2}\left(1 + \frac{BR}{CC}\right)} \tag{1}$$

In (1), the money multiplier $m2$ depends negatively on (i) the public's propensity to hold cash (the cash ratio, $C = CC/M2$), and (ii) the banks' propensity to hold bank reserves (the reserve ratio, $R = BR/CC$). Since the cash and the reserve ratios are easy to calculate, they too can be used for the analysis of the monetary stance of a country, and explain the differences in the money multiplier ratios across countries. For example, the very low level of the money multiplier in Russia and Japan can be partly explained by a high cash ratio. For Russia, a high cash ratio can in turn be explained by a low trust of Russians in the financial institutions and a low level of financial inclusion. The situation in Japan is quite peculiar – in spite of the fact that Japan is a world leader in automation and digitalization, the Japanese have had a long history of cultural preference to hold huge share of their money in a cash form.

Reflecting on the outlined standard approach to the money multiplier, Bain and Howells (2005) state: "This is an account of money supply determination which has proved historically popular and has been kept alive by the popularity of the IS/LM model of a simple macroeconomy. In that model, the LM curve is derived on the assumption of a fixed quantity of money, which is exactly what is provided by the 'monetary base', or 'deposit-multiplier' account of money supply determination."

The notion of multiple deposit expansion remains popular too. A number of explanations may be offered. First, the multiplier mechanism conforms to the conventional wisdom about how banks create money. Perhaps due to the Austrian school, people tend to think that banks are just 'money warehouses' which are unable to create money (which, in turn, is considered to be a commodity, not social relation). According to this simplistic view, banks just redistribute money (commodity by nature) in the economy, lending out idle deposits to gain profit. Under this scheme, the same amount of money belongs to depositors and borrowers at the same time. The activities of banks in redistributing money are assumed to violate this 'monetary law', causing financial crisis in

the end.

Second, the majority of textbooks in economics, money and banking refer to the money multiplier mechanism to explain how money is created. For instance, Samuelson (1948) claims that the 'magic pen' view (money creation 'out of thin air') is erroneous. He states that banks can create money only collectively due to a simple reason: granting loans leads to reserve leakages (from the creditor's bank), so banks tend to lend out virtually the same amount that they previously acquired, and as new loans and deposits circulate, the total amount of money created by the banking sector increases.[3]

Third, multiple deposit expansion is an artefact of macroeconomic models such as Financial Programming Models (FPM) constructed by the IMF for emerging market countries which participate in its stabilization programs. In these models, the money multiplier is the cornerstone of the monetary sector. For instance, in Russia the money multiplier was thoroughly analyzed under monetary targeting at the dawn of the new century. As Ponomarenko et al. (2012) put it, "the Bank of Russia's monetary policy was conducted by setting limits for the growth of the narrow monetary base and other positions of the aggregated central bank's balance sheet in the Monetary Program". Despite shifting to de facto exchange rate targeting in 2000s, the central bank continued to use the analysis of money demand, money velocity and money multiplier dynamics. The Monetary Program is still regularly published in the Monetary Policy Guidelines (full-fledged inflation targeting notwithstanding).

Fourth, still, the money multiplier is frequently mentioned in debates on monetary policy issues in the academic literature. Woodford (2010), for example, describes the Federal Reserve's framework in 2008-2009 in the following way: "Under contemporary institutional arrangements, the Fed automatically adjusts the supply of base money as necessary to maintain its target for the federal funds interest rate; thus, any change in the money multiplier due to a banking crisis should automatically be offset by a corresponding increase in the monetary base, neutralizing any effect on interest rates, inflation, or output".

Figure 1: Money Creation Type 1

Figure 2: Money Creation Type 2

[3] For earlier explorations, see, e.g., Moulton (1918), Phillips (1920), Crick (1927) and Meade (1934); for later developments, see, e.g., Prochnow (1949), Cagan (1965), Humphrey (1987), and Bindseil (2004, 2011).

Since the 1960s, however, the concept of the money multiplier has been facing increasing criticism, mostly in the works of the Post Keynesian school (Minsky 1960; Moore 1979, 1983, 1988; Lavoie 1984; Gedeon 1985; and others) with regard to the nature of modern banking. The Post Keynesians essentially claimed that modern money is endogenous, for the following key reasons.

- Commercial banks create money 'out of thin air', by extending their balance sheets – see Figures 1 and 2. Note, however, the important difference between these figures: while money is created by lending in Figure 1, a domestic (or policy) source of money creation, it is created by the linkage of the home economy, via its external sector, with the rest of the world in Figure 2, a channel of international spillovers that became more prominent with globalization in the recent decades – see, e.g., Cetorelli and Goldberg (2011).

- Modern central banks fulfil commercial banks' demand for reserves, thus softening the supply-side constraints for banks. The chain of causation goes from broad money to monetary base, not the other way round.

- The money multiplier is unstable, it depends on individual portfolio decisions; even being able to control the monetary base, central banks cannot manage broad money.

Outside the Post Keynesian school, Goodhart (1989) – a monetary economist focusing his research on central banking – described the concept of the money multiplier as being so far removed from reality that its appearance in textbooks "amount[s] to misinstruction". The money multiplier approach is also criticized in a few textbooks. For instance, Bain and Howells (2005) note that modern central banks use interest rates, not the monetary base, as their operational targets. As a result, the money supply gains 'additional degrees of freedom', thus becoming less dependent on central bank's decisions. In the 2000s, central bankers followed suit (McLeay et al. 2014; Jakab and Kumhof 2015; Jordan 2018; Borio 2019). Currently, the absence of 'multiplier effects' is a kind of 'internal consensus' among central bankers, especially those involved in monetary policy implementation.

However, it is doubtful that this critique is sufficient to refute the notion of multiple deposit expansion. The argument that "money is created out of thin air, loans create deposits – not the other way round" can rebut only the roughest representations of the process. Other arguments could support the view that now the multiplier does not work, but cannot explain if it could work in the past and why it became so insignificant now. Classical monetary economists, starting with A. Marshall and C. Phillips, admitted that the first stage of the process implies that individual banks create money 'out of thin air'. Yet, in their view, banks faced severe supply constraints which did not let them use their 'money productive capacities' in full (in accordance with their nature as banks). Hence, in-depth critique of the money multiplier should include not only 'out-of-thin-air' rhetoric or the simple description of central bank operations, but also the thorough analysis of supply- and demand-side forces behind money creation. In other words, the combination of supply- and demand-side constraints can create the environment where multiple deposit expansion might be either possible or not. To our knowledge, such a research question has not be explored, and verified empirically, which constitutes the novelty of our present contribution.

Filling in this gap in the monetary economics literature, the aim of this paper is, precisely, to highlight the necessary conditions for multiple deposit expansion to exist institutionally and, furthermore, to test against the data if they are valid for Russia

nowadays. Accordingly, we revisit the issue, first, in a 'narrative' context of the evolving institutional frameworks for banking activities and monetary policy that profoundly and ultimately are shown to shape it out; and, second, empirically too, examining quantitatively the relative importance of money supply and money demand factors on the equilibrium level of real money balances. Using a vector autoregression model estimated on Russian monthly data over two subsamples, 2005-2012 and 2012-2019, we find robust evidence that, while multiple deposit expansion may have existed in underdeveloped financial systems in the past, where the volume of lending was limited by the supply of bank reserves, lending at present is constrained mainly by the demand for credit. We propose the following main explanations of why this has become so: the rapid rise of money markets in the 20th-21st centuries, the unlimited access to central bank liquidity provision facilities, and the evolution of bank management from the 'golden rule' of banking, where liquidity gaps should be zero, to Asset and Liability Management (ALM), where banks should flexibly manage liquidity gaps. Our results robustly show that the influence on real money balances of money supply factors, such as bank reserve requirements and the real monetary base, has become statistically insignificant over the recent decade in Russia, while that of money demand factors, such as the nominal interest rate, has remained significant and negative, which is consistent with the economic intuition we have outlined and expand on further below.

The remainder of the paper is structured as follows. In the next section, we discuss multiple deposit expansion, and how it still matters for underdeveloped monetary systems but not anymore for developed sovereign monetary regimes. Section 3 then outlines the background for our research, namely, the evolution of banking and the stylized facts on monetary policy in Russia. Section 4, in turn, describes our methodological approach, presents the data, reports and interprets the main econometric results, and checks for their robustness. The final section concludes.

2 Multiple Deposit Expansion: Does It (Still) Matter?

Multiple deposit expansion has always been a theoretical model striving to explain how banks operated rather than a collection of empirical evidence. Hence, it was based on a number of explicit or implicit assumptions or implications. They reflected, more or less, banking practices and patterns in monetary systems existing at the time when the notions of 'deposit', 'reserve' and 'money multiplier' were actually being formulated.[4]

2.1 Key Features of Financial Systems of the Past

The founders of the multiplier concept proceeded from the following propositions (Phillips 1920):

1. Banks are subject to supply-side constraints.

    1.1 Bank lending leads to reserve leakages, falling confidence in banks and, as a result, 'bank runs'.[5]

    1.2 Banks cannot effectively cope with these leakages.

[4] Although the idea of multiple deposit expansion appeared in the middle of the 19th century, it was systematically analyzed only after the First World War, by C. Phillips in his "Bank Credit" (1920). He claimed (contrary to the conventional wisdom of the time, credit theory of money, the revival of which we are witnessing nowadays) that individual banks were able to extend only slightly more loans than deposits or reserves they possessed while the banking sector as a whole could multiply deposits or reserves.

[5] The main channels of reserve leakages include external trade (reserve flows followed the circulation of goods, services and capital among countries) and growing demand for cash (as a result of either wage growth or falling confidence).

1.2.1 Banks with scarce reserves cannot borrow them from banks with ample reserves (i.e., the money market is illiquid or even non-existent).

1.2.2 Banks have limited access to central bank facilities.[6]

2. Banks are able to overcome demand-side constraints (stimulate demand for their loans) by widely altering interest rates .[7]

As a result, banks were likely to choose 'cautious' liquidity management strategies. 'Rules of life' for them were developed by commercial loan theory (which is a close relative of the well-known real bills doctrine). The cornerstone was 'the golden rule of banking': "every loan should be matched by a deposit of the same size and maturity" (in modern terminology, banks should maintain zero liquidity gaps). Other rules were just consequences of the latter:

- "Expand lending if reserves grow and curtail should they fall."
- "Provide loans only to industries with rapid circulation of funds – agriculture and trade."

Effectively, this meant exogenous money and endogenous interest rates. Under these circumstances, monetary authorities could prefer trying to manage monetary aggregates and credit by changing reserve requirements or the amounts of bank reserves provided to commercial banks.

2.2 Evolution of Financial Systems

The 20th century saw drastic transformations of financial systems. First, after the collapse of metallic standards external trade ceased to affect automatically bank reserves circulating in the country – in this sense, the banking sector has become a closed loop. Second, after the Great Depression governments and central banks began to smooth out the economic cycle more actively. In particular, deposit insurance schemes became more widespread. This, to some extent, helped decrease the risks of bank runs. Third, the facilities of the central bank in managing the commercial banks became more accessible. As a result, banks' behavior has changed.

Systems of refinancing of commercial banks have undergone significant changes. For a long time, central banks have been following the principles of the lender of last resort, formulated by Walter Bagehot (Humphrey 1989):

- Refinancing is an urgent measure, aimed at preserving stable functioning of the system in whole, not saving particular banks.
- Support may be provided to financially solvent banks facing temporary difficulties, at penalty rates and against 'good' collateral. Financially unsound institutes should go bankrupt.

[6] The idea of limited access to refinancing (due to 'stigma', as made clear next in this footnote) dominated in central banking for almost two centuries. It was developed by 'the father of central banking' H. Thornton and W. Bagehot (see also further below). Both claimed that commercial banks may refer to central bank facilities only under exceptional circumstances ('bank runs') – not regularly, at least – when they lack liquidity but remain financially sound in other respects. In practice, every bank that tried to refer for refinancing was perceived by other banks as financially unsound, and they stopped dealing with it. This type of behavior let some central banks (for instance, the Bank of England) to hold their refinancing rate lower (not higher) than the prevailing interbank rates for years.

[7] "Eastern banking institutions have been able frequently to offer the western borrower lower discount rates than prevailed in the West because of the greater abundance of money in the Eastern banking centers. Banks in New York City, in particular, are said to resort to lower rates as a means of securing new accounts, except in periods of stringent money." (Phillips 1920).

- In 'tranquil' times macroeconomic stability should be maintained through smoothing out the swings of amounts of money in circulation.

Although not having so much in common, these principles let the central bank to fine-tune the system in such a way as to control the amount of reserves. The implementation of these principles prevented the banks from using central bank lending facilities at their initiative. The most powerful disincentives were associated with the so-called 'stigma' we referred to earlier. It was the very cornerstone of the concept of reserve money (or total bank reserves) management.

As financial systems were evolving, it was gradually becoming obvious that the supply constraints on banks are losing effectiveness. Central banks had been struggling to admit this fact for a long time, taking various new administrative measures successfully circumvented by banks. In the end, on the back of the growing popularity of the New Keynesian school of economics, which reduces the monetary sphere to a set of interest rates,[8] regulators have abandoned bank reserves management and started again to influence the economy through interest rates. Operationally this meant a free access to central bank lending facilities for financially sound banks against 'good' collateral.

Banks, in their turn, were gradually moving away from the idea of maintaining zero liquidity gaps. In the 1920s, when the bond market became liquid, banks started to manage the asset side of their balance sheets by creating liquidity buffers (theory of transformation) and matching the loan repayment schedules with the income flows of the borrower (anticipated income theory). In the 1960s, when liquid interbank (uncollateralized) markets appeared, US banks started to manage their liabilities. In the 1990s, banking practices based on asset and liability management (ALM) and transfer pricing (TFP), in other words, liquidity gap management, became widespread. In the 'multiple expansion' era banks had enough space to alter the interest rates on their loans while their credit expansion was constrained by the amount of reserves (interest rates were endogenous, the amount of reserves was exogenous). Now the reverse is true: banks are not liquidity constrained but they cannot steer the interest rates (since nowadays central banks determine the level of interest rates to influence aggregate demand). Are the new methods of liquidity management used by banks riskier, can their use lead to a situation when banks grant loans to every borrower irrespective of their credit quality? It is doubtful. If a bank provides loans to unreliable borrowers, loans are poorly repaid and in the end, the bank faces losses. In practice, in recent decades the control over the creditworthiness of the borrowers was improving. At the same time, credit risk management became separated from liquidity risk management. On the operational level, decisions on bank reserve management and providing a loan are made by different divisions. There is no tight link between them, as it may have been in the past. Now banks first determine the list of creditworthy borrowers and then strive to fulfil their demand for credit, covering the liquidity needs afterwards (if necessary, they sell bonds or borrow in the interbank market or from the central bank). Thus, lending constraints have not disappeared at all; they just have changed their form.

3 Russia: Monetary History in Brief and Some Stylized Facts

We now briefly review the monetary history of Russia and the stylized facts with regard to its banking practices and their evolution.

[8] There are exceptions in this literature, though, where money or real balance effects matter and are explicitly modeled: see, e.g., Leeper and Roush (2003), Ireland (2004), McKnight and Mihailov (2015). On the practical side of central bank policy making, monetary aggregates have also been carefully monitored by the European Central Bank (ECB), ever since it came into existence in 1999, actually inheriting a long German tradition at the Deutsche Bundesbank.

3.1 Evolution of Banking Practices in Russia

In Russia, the road from the 'multiplier practices' to ALM has taken approximately 25 years. In the episodes of tight monetary policy in the 1990s, the monetary authorities tried to curb inflationary pressures through the contraction of the money supply, refusing to provide the necessary reserves to banks. The money market faced a number of crises. In this environment, the majority of banks by the beginning of the 21st century had to employ conservative liquidity management practices. In the 2000s, rising oil prices and exchange rate targeting by the central bank accompanied by massive foreign exchange (FX) interventions created a liquidity surplus. Banks, feeling comfortable with loose liquidity conditions, switched to riskier lending practices. In the era of ample liquidity and rapid economic growth there was no need for the ALM concept. The situation has changed dramatically after the GFC, as in 2009 Russian GDP shrank by 8%, and around $250 bln or 40% of international reserves were lost. Banks faced with the materialization of both exchange rate and liquidity risks. This may have forced the biggest players in the market to implement the frontier risk practices of leading banks in Europe and the US. As a result, ALM has appeared in Russia. The changes in monetary policy, namely, the transition to a floating exchange rate and the steering of interest rates, were also supportive to the establishment of the modern way to influence the banking sector by the central bank – through the demand for credit, not the supply of reserves.

3.2 Stylized Facts on Money and Banking in Russia

To illustrate the absence of multiple deposit expansion in Russia nowadays, we will provide some evidence on the characteristics of the banking system (the conditions for multiple deposit expansion were noted earlier). We will mainly focus on the period from 2005 to 2019.

Before 2004, retail depositors had no guarantee that they could get their money back should the bank go bankrupt. The probability of 'bank runs' started to decline in 2004, as the Deposit Insurance Agency (DIA) was established. Prior to 2004, the share of cash in M2 was hovering between 35% and 40%; after the DIA was established, the ratio declined more than twofold – see Figure 3.

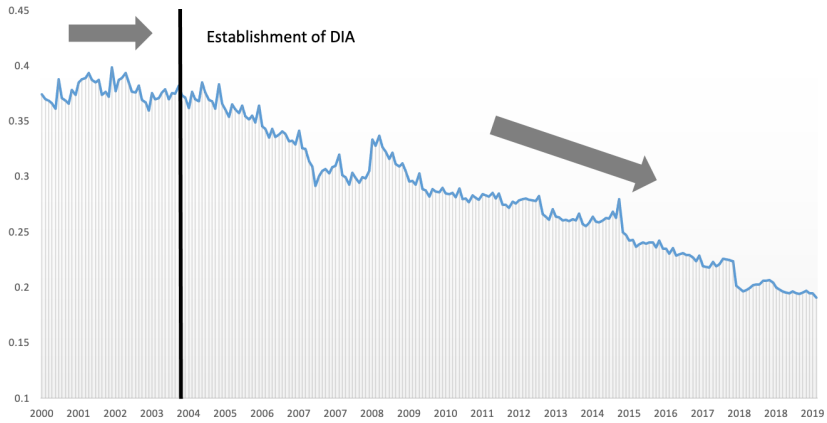


Figure 3: Share of Cash in Circulation in M2 in Russia. Source: Bank of Russia

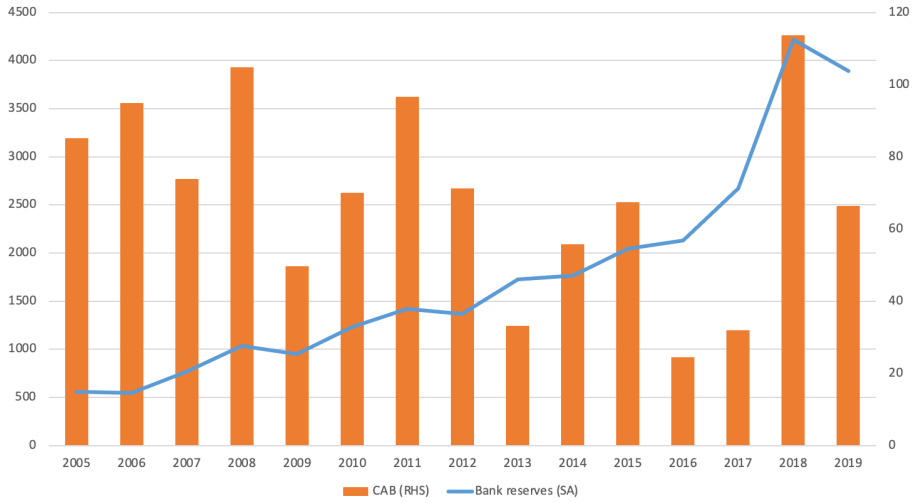


Figure 4: Bank Reserves and Current Account Balance Dynamics. Source: Bank of Russia

In the period under consideration bank reserve dynamics was not driven by external trade, and the correlation between the two series is even negative – see Figure 4. Although the upper boundary of the interest corridor was finally introduced only in 2007, interest rates before 2007 could not be regarded truly endogenous – see Figure 5. Generally, this was a period of ample liquidity (owing to massive purchases of foreign currency by the central bank). Hence, interest rates fluctuated mainly close to the lower boundary (the deposit rate). In times of growing liquidity needs (for example, in tax periods) rates went up. In these cases, the Bank of Russia held repo auctions to provide liquidity. Nevertheless, the amplitude of interest rate fluctuations was rather high (by the standards of developed countries) – about 300-400 basis points, as Figure 5 shows.

During its transition to inflation targeting (introduced in 2015), the Bank of Russia was enhancing interest rate management and reducing its interventions in the foreign exchange market. As a result, at the beginning of the 2010s the Bank of Russia was already able to manage interest rates in a narrow corridor – see again Figure 5. Interest rates could be regarded exogenous (in the sense that the central bank, not commercial banks, determines them).

In crisis episodes interest rates, however, rocketed, going far beyond the corridor. This was conditioned by severe lack of collateral (and, thus, inability of banks to refer to central bank refinancing facilities). To mitigate the liquidity needs, the central bank in those circumstances eased collateral requirements, granted a bulk of loans against non-marketable assets (in 2008-2009 and 2014) or even provided unsecured loans (in 2008-2009). In effect, banks could feel themselves in a more secure position compared to banks under the gold standard that could not refer to central bank refinancing freely.

We can summarize the outlined institutional trends and the criteria for the money multiplier to operate at the background of their evidence in Russia as in Table 2. The main conclusion is that, in whole, the Russian evidence offers a different explanation of money multiplier dynamics.

The GFC puts the multiple deposit expansion doctrine into doubt. According to the doctrine: (i) the multiplier is stable – since its components (the share of cash in money supply and the proportion of bank reserves to cash) remain on a steady path, at least

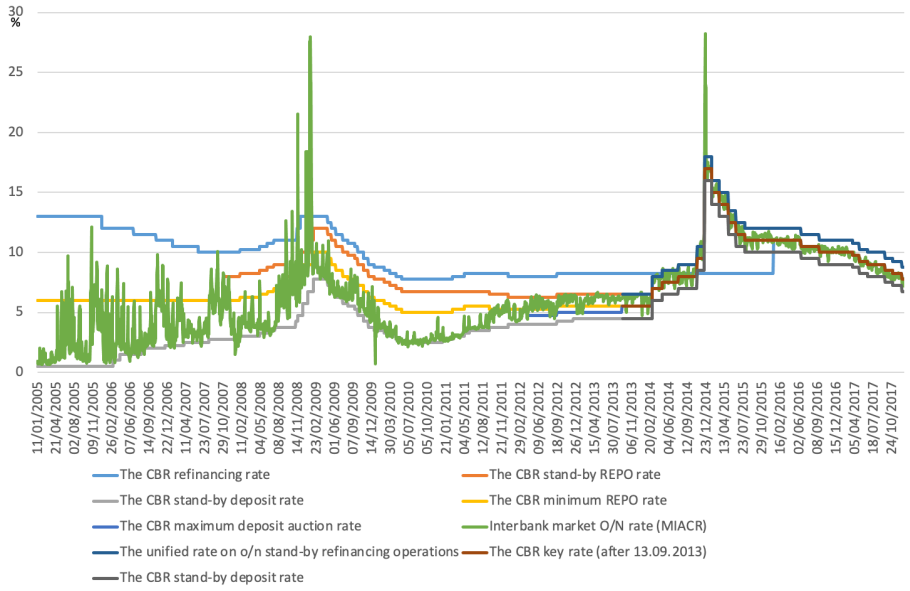


Figure 5: Bank of Russia Operational Framework Evolution. Source: Bank of Russia; MIACR is the interest rate on overnight interbank loans

in the long run; (ii) in times of crisis the multiplier may somewhat go down (due to weak demand combined with rising willingness of the public to possess cash), but with an inevitable offsetting pick-up (i.e., faster money supply growth) after the slump ends.

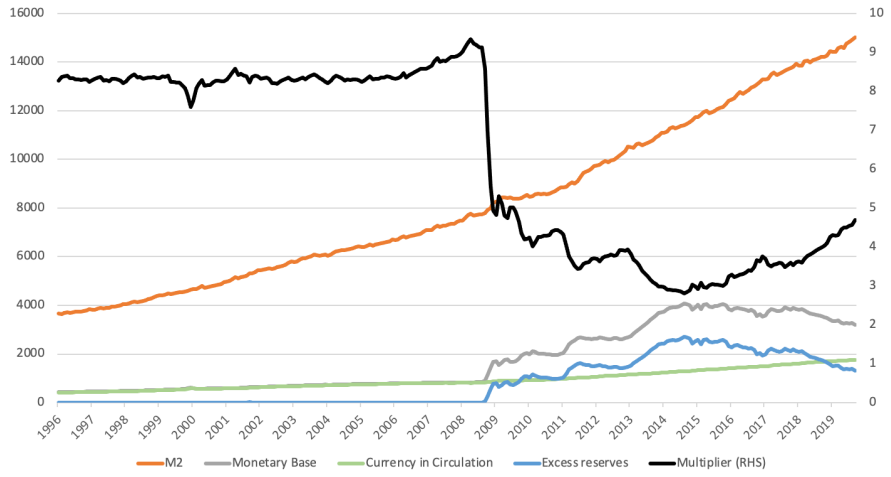


Figure 6: Monetary Variables in the USA. Source: St. Louis Fed.

In short, neither of these predictions came true. In the USA, the collapse of the multiplier – see Figure 6 – was mainly driven by a massive rise of bank reserves (not a propensity to hold cash!) as a result of the Fed's liquidity injections. Moreover, in 2017-2018, when the economy was growing faster than 2%, the inflation threat also failed to materialize. The rebound in the values of the multiplier since that period can be explained not by enhancing lending opportunities for banks but by the 'quantitative

#	Criterion	Evidence for Russia
1	Banks are subject to supply-side constraints	Virtually NO (only in times of crisis)
1.1	Bank lending leads to reserve leakages, falling confidence in banks and, as a result, 'bank runs'	NO (DIA prevents 'bank runs', reserves do not depend on external trade)
1.2	Banks cannot effectively cope with these leakages	Virtually NO (only in times of crisis)
1.2.1	Banks with scarce reserves cannot borrow them from banks with ample reserves (money market is illiquid or even non-existent)	Virtually NO (money market is usually liquid)
1.2.2	Banks have limited access to central bank facilities	Virtually NO (banks could borrow freely against good collateral)
2	Banks are able to overcome demands-side constraints (stimulate demand for their loans) by widely altering interest rates	Virtually NO (central bank more or less controls interest rates)

Source: compiled by the authors.

Table 2: Criteria of Multiple Deposit Expansion and Russian Evidence

tightening' (balance sheet contraction conducted by the Fed) – at the time, the total assets shrank from $4.4 trn to $3.8 trn.

Some authors tended to ascribe the increased rigidity of broad money to monetary base dynamics in 2007-2009 by institutional developments in the US banking sector. In the last quarter of the 20th century these tendencies manifested themselves in the growing role of non-bank liabilities and non-reservable sources of funding for banks (see, e.g., Carpenter and Demiralp 2012). While admitting this kind of explanation from an empirical perspective, we note that it does not contradict the proposition of multiple expansion of reservable liabilities (which we reject as well; the hypothesis can be tested for countries with high share of reservable liabilities in the broad money such as Russia[9]).

Both in 2007-2009 and 2020 broad money growth was affected by the actions of the Fed, but not in the way the multiplier concept implies. As McLeay et al. (2014) show, quantitative easing (QE) programs lead to broad money growth since banks act as intermediaries in the asset purchases: in the end, the central bank buys papers from non-banks in exchange for money. This yields nearly a 1:1 proportion between QE and broad money growth if asset purchases are large enough.

Finally, the dynamics of the money multipliers in the US and Russia – see Figure 7 – can provide only scant information about the ability of commercial banks to extend loans on the base of existing reserves. For instance, up to the GFC, bank reserves in the US mainly (up to 99%) consisted of required reserves since the Federal Reserve was successful in banking sector liquidity management (by conducting operations every day and creating strong incentives for banks to get rid of excess reserves – they were not remunerated at the time). The Bank of Russia in the same period managed the exchange rate, contributing to a liquidity surplus and enabling interest rates to be more volatile. Under these circumstances, Russian banks had to maintain more reserves due to the higher uncertainty of liquidity conditions. In 2013, as the transition to inflation targeting (IT) in Russia gained momentum, a symmetric 200-basis-point interest rate corridor was formed, and the central bank enhanced the banking sector liquidity management under a structural liquidity deficit. As a result, excess reserves became less needed. Figures 8 and 9 provide a comparison between bank reserves structures and trends in the US vs Russia.

[9] See Bank of Russia's Regulation #507-P (ed. 30.08.2020) "On required reserves of credit organizations".

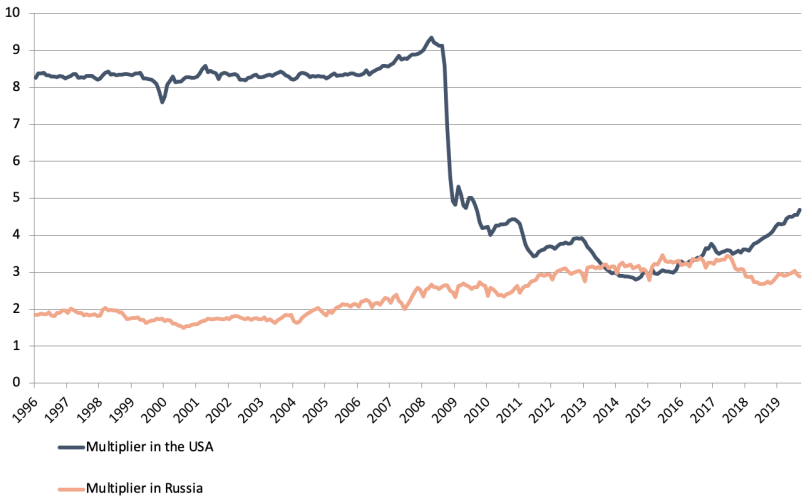


Figure 7: Money Multipliers in Russia and the USA. Source: St. Louis Fed.

Comparing the balance sheet structures of a typical bank in the 20th century following the 'golden rule of banking' and that of the biggest modern Russian bank, Sberbank (see Figure 10), we can conclude that bank balance sheet structure has dramatically changed. Instead of maintaining zero liquidity gaps, as in the past, a commercial bank nowadays in Russia manages them.

4 An Empirical Investigation of Money Creation in Russia

Money creation in Russia has been analyzed by other authors from two main angles. The first strand of papers considers money creation as a result of domestic banks' activities – i.e., as reflected by money creation type 1 in Figure 1. Our study falls into this category. The second one puts the process in a broader context, taking into account the influence of distinctive features of emerging markets, such as FX interventions, volatile trade and capital flows, and dollarization – i.e., as reflected by money creation type 2 in Figure 2.

The first group of papers includes Vymyatnina (2006, 2013), Badarudin et al. (2009), Juurikkala et al. (2011) and Salmanov et al. (2015). Vymyatnina (2006, 2015) finds Granger and Sims causality from credit to money for 1995-2004 and 1995-2011, concluding that money in those periods was endogenous. Badarudin et al. (2009), however, get the reverse result, having used the same methodology. We would explain the huge difference in these causality results by discrepancies between Russian statistics (Bank of Russia and Rosstat data; used by Vymyatnina) and international ones (IMF International Financial Statistics Database, used by Badarudin et al.) We, in our turn, expand their analysis and make it more precise methodologically and quantitatively by using not only causality tests, as a preliminary pass, but also considering impulse responses and variance decompositions from a vector autoregressive (VAR) empirical model featuring the key factors of money demand and money supply. Juurikkala et al. (2011) followed another, micro-based approach (panel regressions), and found the existence of a bank lending channel in Russia for 1999-2007. This is broadly in line with our finding of significant changes in banks' behavior in Russia after the GFC, to be discussed in more

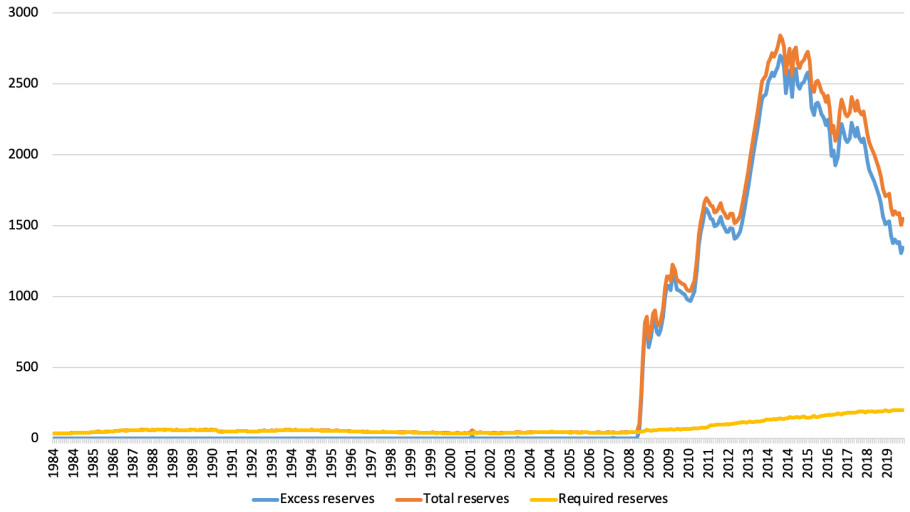


Figure 8: Bank Reserves Structure in the USA. Source: St. Louis Fed.

detail in what follows. Salmanov et al. (2015) claim to have revealed the causal link between the monetary base and the broad money aggregate for 2002-2013 using linear regressions. However, their findings are questionable due to their failure to account for the role of bank reserves (they confuse monetary base with the M0 aggregate).

The second group of studies on money creation in Russia comprises Ponomarenko (2017) and Ponomarenko (2019). Importantly, Ponomarenko (2017, 2019) finds that foreign exchange interventions by the Bank of Russia made a significant contribution to broad money growth in 2001-2014. It is worth noting that he used the IMF's definition of broad money (M2X which includes both ruble and US dollar accounts). As a result, he tends to interpret money that had been actually created by non-resident banks to buy Russian exports as a source of broad money in Russia. At the same time, US dollars cannot be used for payments in Russia, and, hence, M2X has weaker correlation with aggregate demand. In our opinion, broad money in the national definition (M2) is a more appropriate variable for other domestic macroeconomic indicators if we want to focus on domestic credit as a source of money creation.

To support the conclusions drawn from our narrative analysis of the institutional evolution of the Russian banking sector thus far, we next build an empirical model of money creation and estimate it for Russia. The methodology used is a standard VAR approach for policy analysis, and has been applied recently by Carpenter and Demiralp (2010) to study the US case. We follow the macroeconomic dimension of their approach to also enable comparability between the two similar investigations. As we mentioned in the introduction, such a quantitative empirical literature to check whether the money multiplier theory is still relevant to modern banking systems and monetary policy frameworks is only nascent, which increases the value of our current contribution.

4.1 Methodological Approach

To evaluate the impact of supply and demand forces on money dynamics in Russia, we employ a vector autoregression model[10] – VAR(y_t, x_t) with p lags.[11] This methodological

[10] Below we discuss the algorithm of estimation of VAR which consists only of endogenous variables (y_t). Adding exogenous variables (x_t), as we do in our empirical implementation on Russia (to be discussed soon), does not affect the essence of the process.

[11] Our data, code and detailed estimation results are available upon request.

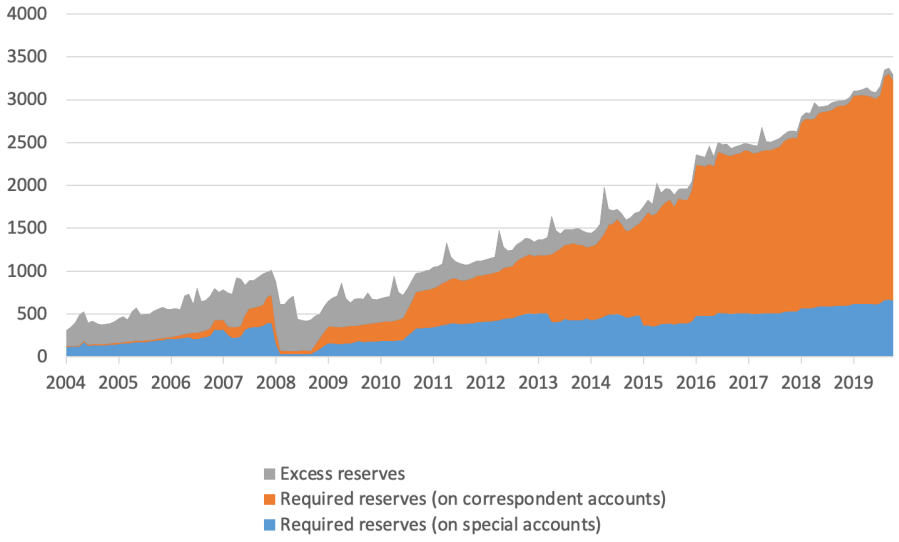
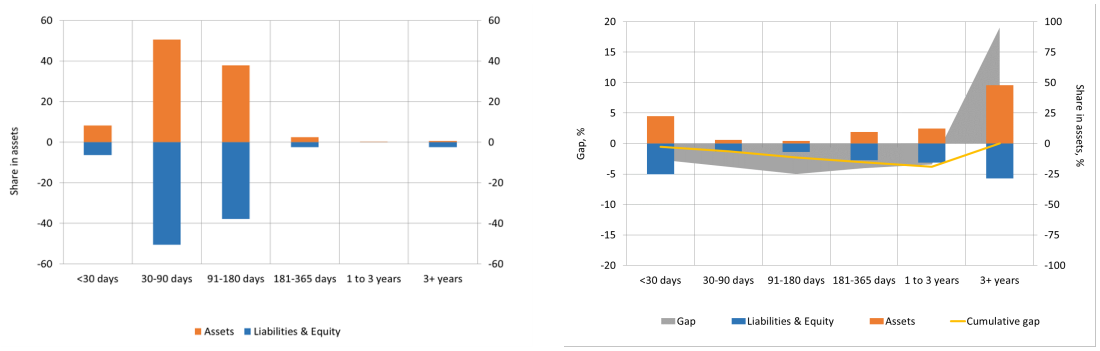


Figure 9: Bank Reserves Structure in Russia. Source: St. Louis Fed.



Stylized example, beginning of the 20th C: gap is always close to zero ('golden rule of banking')

Sberbank, Russia, 2017 (ALM): for short maturities gap is negative, for longer maturities it is positive (emergence of mortgages and other products)

Source: kuap.ru/

Figure 10: Liquidity Gaps – Past and Present

approach generally follows the tradition of Sims (1980), as well as Bernanke and Blinder (1992) and Carpenter and Demiralp (2012) when applied more directly to the issues of interest here. Its major benefit has been viewed in relying strongly on the data – rather than on limited, evolving or controversial theories – to summarize the relationships among the key variables selected in the VAR model. Differently from Bernanke and Blinder (1992) and Carpenter and Demiralp (2012), who use six endogenous variables with monthly US data and six lags in their main specification, we allow for an exogenous variable (which is a measure of the output gap and is meant to reflect the state of the macroeconomy), and work with four monetary and banking endogenous variables, in effect specifying a more parsimonious model with two lags in two subsample periods.[12] A

[12] We test for structural breaks (more detail will follow), and find empirical support for a structural change in early 2012 – as was discussed already in our analysis of the evolution of the institutional framework for monetary policy and banking activities in Russia. We also test for lag length of up to 13 lags, given that we work with monthly data, and use 5 alternative lag length criteria to judge: these are, however, quite diverging for the first subsample, 2005:01–2012:02, two of them selecting lag of 1, another two lag of 3, and one criterion selecting lag of 12. In the second subsample, 2012:03–2019:12,

second difference from Carpenter and Demiralp (2012) is that they use log-levels in their VAR, which may have impaired their findings; whereas we check carefully for stationarity using the ADF and KPSS tests that are based on an alternative null hypothesis, and accordingly enter the nonstationary variables, real money balances and the nominal monetary base in first log-diffrences, i.e., percentage changes.[13] The model is estimated on Russian monthly data (all seasonally adjusted at source, except bank reserves that we also adjusted seasonally via the standard X-13 TRAMO-SEATS procedure), but we do compare our results to the US benchmark, as we have done in some of the earlier figures related to monetary and banking trends and patterns in both these major economies.

To be more precise and following the standard methodology, we estimate the VAR model in its reduced form:

$$y_t = \mathbf{B}_0 + \mathbf{B}_1 y_{t-1} + \mathbf{B}_2 y_{t-2} + \cdots + \mathbf{B}_p y_{t-p} + v_t \tag{2}$$

where y_t is a vector of endogenous variables, $\mathbf{B}_p$ are coefficient matrices, and v_t is a vector of estimated residuals (errors) of the model.

The obtained coefficients may only provide conclusions about the stability of the model and the appropriate number of lags – since the residuals of the model are correlated. To interpret them, we transform the model, getting the structural representation of the VAR:

$$\mathbf{A}_0 y_t = \delta_0 + \mathbf{A}_1 y_{t-1} + \mathbf{A}_2 y_{t-2} + \cdots + \mathbf{A}_p y_{t-p} + \varepsilon_t \tag{3}$$

where $\mathbf{A}_p$ and δ_0 are coefficient matrices, and ε_t is a vector of structural (non-correlated) shocks (disturbances) of the model.

The process of transformation of (2) into (3) is known as VAR identification. We use the simplest and the most popular method of identification – Cholesky identification, according to which the sequence of influence of some variables on others may be determined via causality tests, i.e., by the observed data, or based on minimal and core economic theory, i.e., by imposing economic logic (which, in effect, we do in the Cholesky ordering described further below in two alternatives, for robustness).

4.2 Data

While choosing the appropriate variables, we proceeded from the following guiding principles, pertinent to VARs. First, a model of this kind should consist of a few variables, since VARs are subject to the 'curse of dimensionality'. Second, the choice of variables ought to be justified in terms of monetary theory. There are a number of money demand theories (those of Keynes, Hicks, Baumol, Tobin, or Friedman). We will focus on the 'core' assumptions common to all these theories. Simply put, aggregate demand for money is dependent on its opportunity cost (the interest rate, I) as compared with other (interest-bearing) assets, as well as on aggregate income (Y) and the general price level (P):

$$M^D = f(I, Y, P) \tag{4}$$

Neoclassical, as well as Keynesian, theories of money supply conform well with the concept of the money multiplier. Advocates of both strands of thought agree that money

lag selection is unanimous by the 5 criteria, always selecting lag of 1. For the sake of parsimony and comparability across the two subsamples, we choose to work with lag of 2 in both of them.

[13] The reported VAR estimation in Bernanke and Blinder (1992) is also in log-levels, and in real terms for the three bank balance-sheet variables they use, namely, deposits, securities and loans. However, they state in their footnote 41, p. 918, that in alternative regressions, the balance-sheet variables were taken in nominal terms but "[t]his made little difference. Results were also similar when we differenced the data instead of using levels."

supply depends on the monetary base (B) and the required reserve ratio (R), since all formulas of the money multiplier have this ratio in the denominator:

$$M^S = g(B, R) \tag{5}$$

We should stress here that both money supply and the demand for money are non-observed variables. Statistics shows only the result of their interaction over time. By estimating our model, we will try to decompose the dynamics of money into demand- and supply-side forces.

Having in mind the constraints noted above, we build a VAR(y_t, x_t) consisting of four endogenous and one exogenous variables: $y_t^T = \{B, I, R, M\}$, $x_t = \{y\}$, where: $B \equiv \Delta \ln B$ and R are supply factors (the nominal monetary base and the required reserve ratio, respectively); I is a demand factor (short-term nominal interest rates, in % per annum); $M \equiv \Delta \ln M$ (real money balances, deflated by the CPI) is the result of the interaction of real money demand and supply (lending or a monetary aggregate); y is a measure of the output gap obtained via a Hodrick-Prescott (HP) detrending of the output volume index of goods and services of core economic activities (estimated and seasonally adjusted by the Development Center of the Higher School of Economics), in % deviation from the HP trend. Note that, in effect, all five time series in our VAR model enter in % changes, and have been confirmed to be stationary in this transformation.

According to the logic of the money multiplier, an increase in the monetary base either due to reduction in the required reserve ratio or as a result of operations of commercial banks with the central bank leads to an expansion of deposits to such an extent in which the relation of broad money to the monetary base remains constant. This is explained by the fact that both banks and the non-bank private sector agents strive to maintain reserve-to-deposit and cash-to-deposit ratios stable. As a result, assuming the logic of the money multiplier, we would expect a positive impulse response of real money balance growth and a negative impulse response of the interest rate to shocks to the required reserve ratio or to (real) monetary base growth.[14]

The monetary base consists of currency in circulation (including that in ATMs and banks) and bank reserves. In the modern world, currency in circulation is usually a passive component of the monetary base which changes in response to broad money dynamics, at least, for two reasons. First, money is created by banks through lending or asset purchases or, alternatively, by fiscal authorities by crediting recipients' accounts. Both ways imply non-cash settlement. True, central banks may alter the amount of money in circulation directly through their operations with non-banks. In Russia it may have been the case in the 2000s, in times of massive interventions in the foreign exchange market. Now the amounts of cash central bank operations are negligible. Second, the propensity of non-banks to hold cash may affect the money creation process but cannot be regarded as its main driver.

We use a broad monetary base concept which includes three types of bank reserves – see again Figure 9. The first one is funds in correspondent accounts at the central bank. This is the most liquid part of bank reserves which banks use for effecting payments. At the same time, part of these funds are required reserves which banks have access to (averaged reserves – as opposed to funds in special accounts – see below). Rising correspondent accounts may indicate, according to the concept of the money multiplier, a potential lending growth, while rising required reserves maintained in correspondent accounts point to a lending growth which has already taken place.

[14] According to the Classical Quantity Theory of Money, in the long run real money balances remain constant while in the short run they may fluctuate in response to monetary base expansion. Instead, J.M. Keynes argued that the growth in real money balances may be more protracted, assuming lack of unemployment, liquidity trap or weak reaction of investment to interest rates (Humphrey 1974). Anyway, a positive reaction of a growth in real money balances to a rise in monetary base growth is compatible with both opinions.

The second component of bank reserves is an increase in special (required reserves) accounts at the central bank. Funds are added to these accounts when banks transfer (to fulfil reserve requirements) these funds from correspondent accounts. Funds in special accounts are 'temporarily frozen' for the whole averaging period. At the same time, we include into the model these 'frozen' reserves because they reflect the results of lending already provided. The money multiplier logic implies that as lending grows, the structure of bank reserves changes – excess reserves in the correspondent accounts gradually turn into reserves 'frozen' in special accounts or maintained in the correspondent accounts throughout the averaging period to fulfil reserve requirements.

The third component of bank reserves is funds temporarily absorbed by the central bank using its operations. In other words, these are funds in deposit accounts of the central bank and funds paid for central bank bonds. Banks have different access to all these funds. Funds may fall into overnight deposit accounts at any moment of the day and then return to correspondent accounts of banks next day in the morning (thus, banks have access to them all the time). Likewise, banks lose access to funds in one-week deposits almost for a week, and to funds paid for central bank bonds for several months (individual banks may sell their bonds, but this action does not alter the banking sector liquidity position against the central bank). We do not have statistics on intraday reserve movements between banks, and cannot find out when funds fall in deposit accounts with the central bank. Thus, we include into bank reserves deposits with the central bank (as those that may be 'used' for future lending) but exclude central bank bonds.

We also use the following data for the endogenous variables in our VAR model: reserve requirement ratios (at the end of each month) and short-term interest rates on bank loans of the 30 largest banks in Russia to non-financial organizations for a period of less than 1 year (monthly weighted averages). As an exogenous variable in our VAR model we use the output gap (proxied by the cyclical component of the index of production of base industries extracted by HP-filtering). We proxy real money balances as a ratio of broad money (M2) to the overall price level (measured by the CPI). Where the raw data are not seasonally adjusted at the source of statistical information, we use seasonal adjustment. The time period for all variables is January 2005 – December 2019. Descriptive statistics for our key variables of interest entering the VAR estimation are collected in Table 3.

	R	B	I	M	y
Mean	3.9	82.0	11.6	0.7	0
Median	4.3	71.5	10.6	0.9	0
Max	5.5	750.5	20.0	3.9	7.8
Min	0.5	–672.3	7.9	–8.4	–7.0
SD	1.0	208.3	2.6	1.6	2.8
Obs	179	179	179	179	179

Notation and measurement:
R: reserve requirement ratio, % of total bank liabilities to non-financial corporations
B: nominal monetary base (in log-differences), bln RUB
I: short-term nominal lending rate to non-financial corporations, % per annum
M: real money balances (in log-differences), M2 SA divided by CPI, bln RUB
y: output gap estimated with Hodrick-Prescott filter, % deviation from trend

Table 3: Descriptive Statistics for Russian Monetary Policy Indicators

Multiple breakpoint tests have revealed possible break dates: October 2008, April 2011 and February 2012. The first one may be connected to the GFC and may be included as a dummy, which we did; the latest two, however, may be regarded as sample-splitters. The reason is that in 2011-2012 the Russian banking system was gradually

(starting from the biggest banks and then spreading to the rest of the banking sector) shifting from a structural liquidity surplus to a structural liquidity deficit – see Figure 11. That shift was critical for banks since it spurred them to manage their own liquidity using modern ALM tools (which were unnecessary in the ample liquidity conditions that prevailed before 2011-2012). The precise borders of that process cannot be inferred since they depend on our understanding of what the structural liquidity position is. The Bank of Russia's data shows the beginning of the deficit in late 2011 but we regard February 2012 as a sample-splitter because starting from that point the liquidity deficit became binding for the majority of banks.

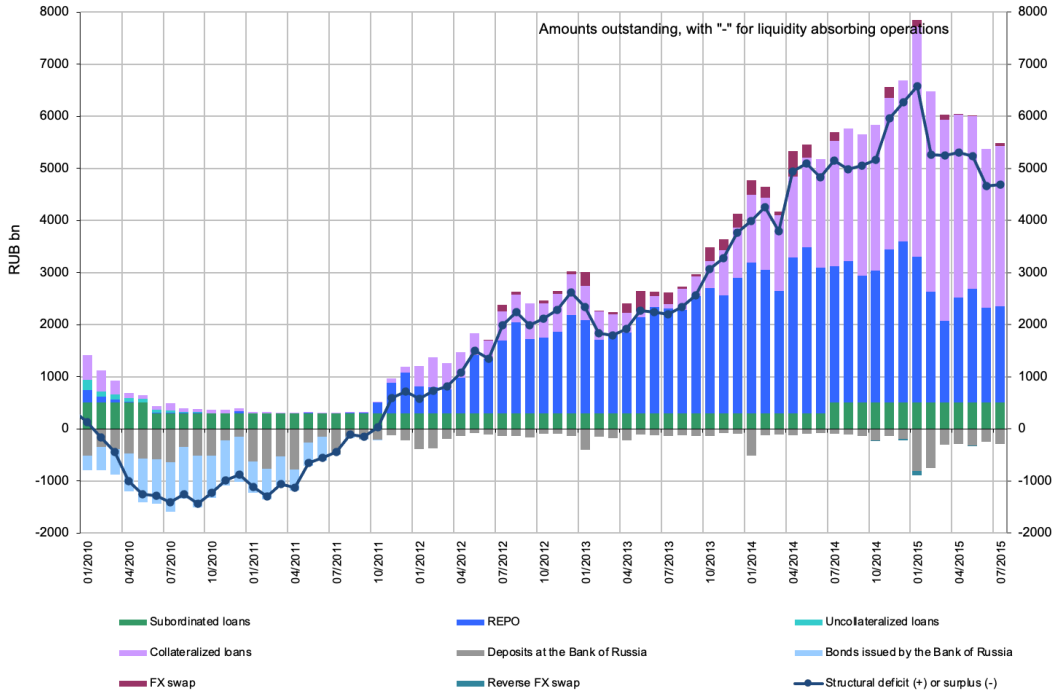


Figure 11: Evolution of the Structural Liquidity Position of the Banking System in Russia, 2010–2015

We take this structural break into account by dividing the whole sample into two parts: January 2005 – February 2012 and March 2012 – December 2019. From the perspective of our analysis of the institutional evolution of the banking system and monetary policy, in particular in the case of Russia we study here (comparing to the US), we expect to uncover econometrically a more prominent role of demand factors in the second period, where Russian banks had to employ modern liquidity management strategies; while in the first subsample period multiplier practices may have played some role, in the second one they should have been inadequate as an explanation for the money creation process.

While making Cholesky identification, we use the following 'core' theory-grounded ordering of the variables: 'required reserve ratio → monetary base → interest rate → real money balances'. The reason is the following: according to the money multiplier logic, at the first stage the central bank decreases reserve requirements (supply-side factor). This leads to bank reserve (and monetary base) expansion (supply-side factor), then rising reserves lead to declining interest rates (demand-side factor), and, finally, banks grant more or less loans (as a result of interaction of supply- and demand-side factors).

Augmented Dickey-Fuller (ADF) tests pointed to probable unit roots in the interest rate and reserve requirements series. However, since this makes little economic sense and contradicts with the results of Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests,

we consider these time series stationary. The other time series we use in the VAR, appropriately transformed, are also stationary, according to these two tests.

4.3 Estimation Results

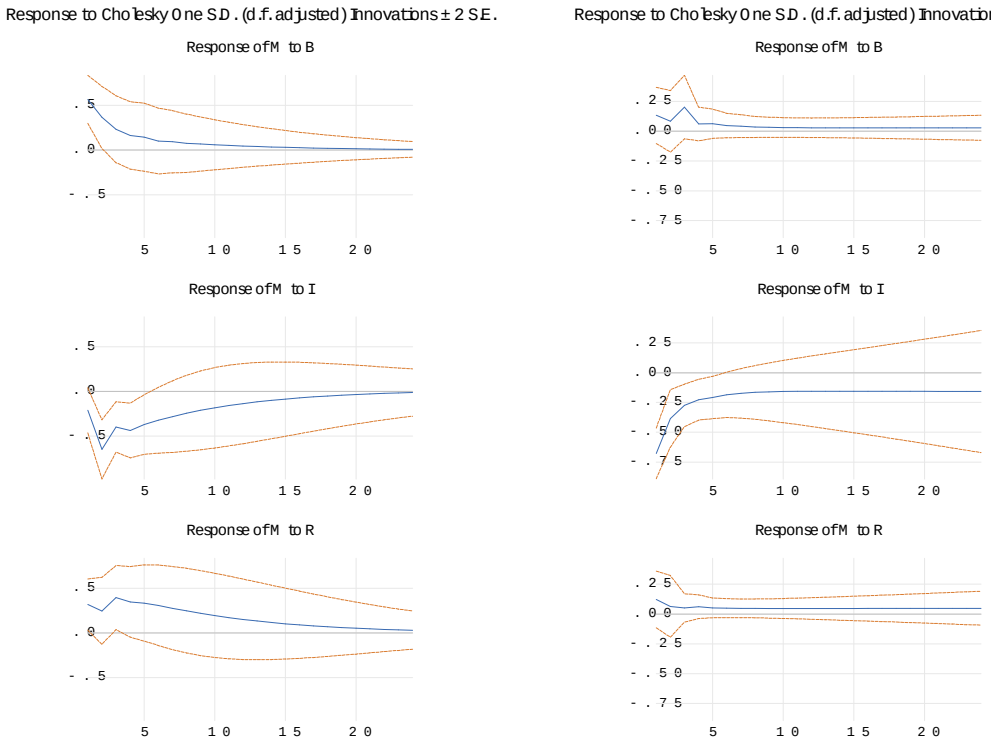
To find out which factors determine the dynamics of real money, we first run Granger-causality tests and then study impulse responses of real money to shocks to the supply-side factors in the model (the monetary base and the required reserve ratio) and the demand-side factor (the interest rate). Pairwise Granger-causality tests shows that at a 10% confidence level only real money M and the interest rate I Granger-cause each other. At the same time, real money also Granger-causes reserve requirements R and the monetary base B (see the full-sample column in Table 4). Our Granger-causality test results supports the notion that neither reserve requirements nor the monetary base determine real money dynamics. This general finding is also true if we look by subsample (see the 1st and 2nd subsample columns in Table 4): again, reserve requirements R and the monetary base B do not Granger-cause real money balances M; whereas M used to Granger cause both R and B in the first subsample, but lost this predictive power in the second subsample.

H_0	p-value		
	full sample	1st subsample	2nd subsample
B does not Granger cause R	0.1498	0.0748	0.4233
R does not Granger cause B	0.1501	0.4633	0.0771
I does not Granger cause R	0.3057	0.0038	0.4046
R does not Granger cause I	0.3408	0.0216	0.0626
M does not Granger cause R	0.0759	0.0187	0.8794
R does not Granger cause M	0.9325	0.6115	0.1409
I does not Granger cause B	0.3490	0.4884	0.3834
B does not Granger cause I	0.4527	0.0979	0.9678
M does not Granger cause B	0.0043	0.0005	0.3178
B does not Granger cause M	0.8455	0.7587	0.6375
M does not Granger cause I	0.0181	0.1612	0.0477
I does not Granger cause M	0.0659	0.1109	0.2338
	2005:01–2019:12	2005:01–2012:02	2012:03–2019:12
	178 or 179 obs	84 or 85 obs	93 obs

Table 4: Pairwise Granger Causality Tests for Russian Monetary Policy Indicators

Figure 12 illustrates the impulse responses from our estimated monthly VAR model over two subsamples: 2005:01-2012:02, in the left-hand side panel, and 2012:03–2019:12, in the right-hand side panel.[15] It shows empirically, and consistent with our theory-grounded and institutionally-interpreted predictions earlier, that in the first subsample period shocks to supply-side factors (the monetary base B and the reserve requirement ratio R) affect significantly and positively (over about 2-3 months) real money balances; however, in the second subsample period, statistical significance is lost, on both accounts. By contrast, shocks to demand-side factors (the nominal interest rate I) have a significant negative impact (over about half a year) on real money balances, with the effect somewhat stronger in absolute magnitude (according to the path of the point

[15] Figures 14 and 15 provide a more detailed view, involving all impulse responses from the VAR by our two subsamples, respectively.



First subsample: 2005:01-2012:02 Second subsample: 2012:03-2019:12

Figure 12: The Impact of Supply- and Demand-Side Factors on Lending

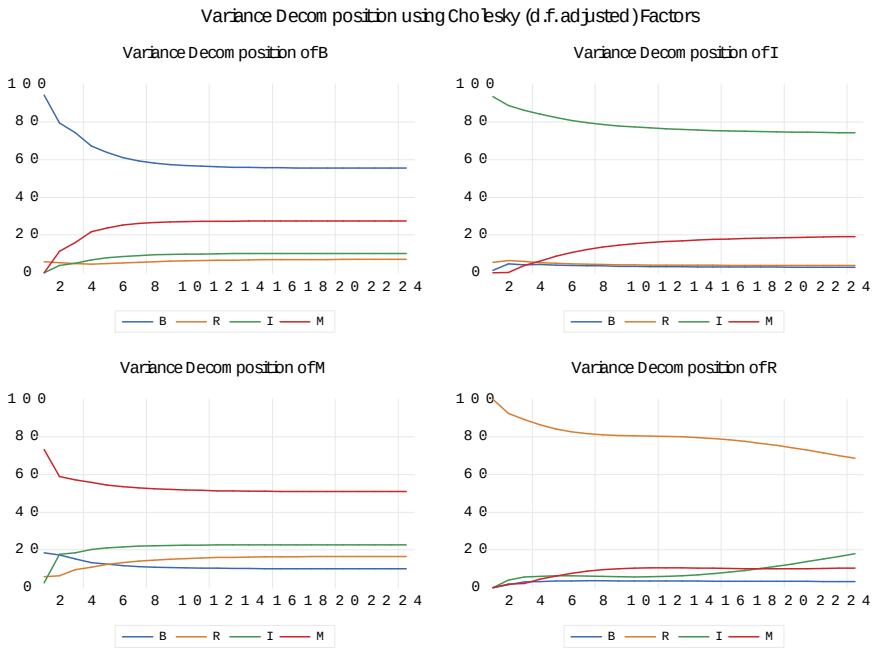
estimate) in the second subsample period. We can, therefore, conclude that while before 2012 the behavior of Russian banks has been broadly consistent with the money multiplier logic, after 2012 this has been no longer true (according to our econometric estimation). That is, currently we do not find empirical evidence of multiple deposit expansion in Russia, i.e., the money multiplier does not appear to function (according to our empirical assessment) in the Russian banking sector. Forecast error variance decompositions from our estimated VAR model, by the same two subsamples (see the two respective panels in Figure 13), further support convincingly these key takeaways from our work.

4.4 Robustness Tests

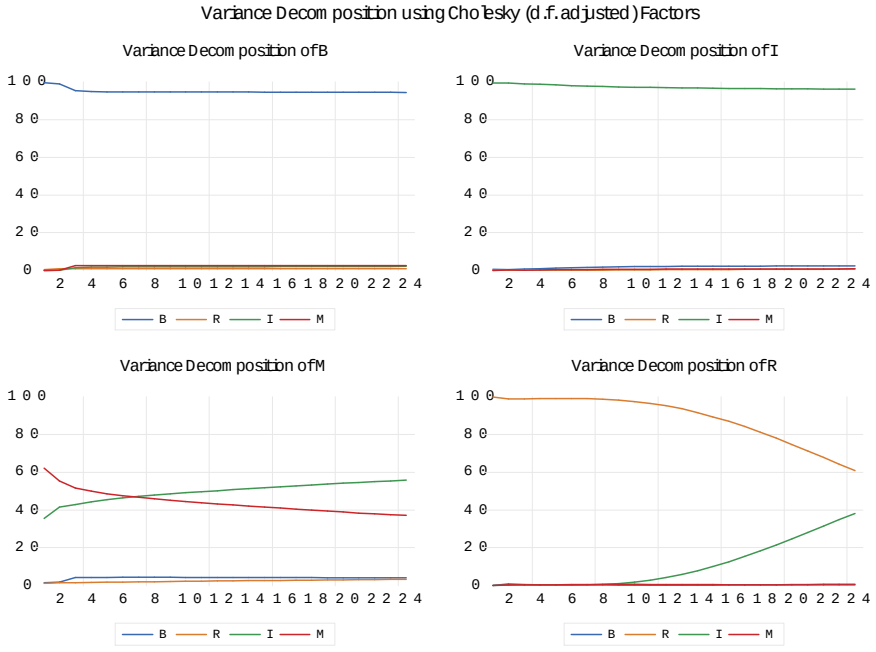
To check the robustness of the reported main results, we considered other schemes of identification of our core VAR model.

1. Another order of the endogenous variables in the Cholesky decomposition. An order 'required reserve ratio → monetary base → real money balances → interest rate' may also be appropriate from a theoretical as well as operational perspective, as we argued earlier. Nevertheless, the shape of impulse responses does not change in any essential way (see figures 16 and 17).

2. Generalized impulse responses. This approach is due to Pesaran and Shin (1998) and constitutes an alternative to the traditional one of orthogonalized impulse responses. As noted, e.g., in Mihailov (2009), its main advantage is that it does not require orthogonalization of shocks and is, thus, invariant to the ordering of variables; yet, there is a disadvantage too, in that it requires an assumption on



First subsample: 2005:01-2012:02



Second subsample: 2012:03-2019:12

Figure 13: Forecast Error Variance Decompositions

the distribution of shocks (usually, multivariate normal or the empirical distribution observed historically). Again, the results remain quantitatively as well as qualitatively unchanged (see figures 18 and 19).

Thus, both alternative methods of VAR identification confirm the results we reported and discussed in the present section.

5 Concluding Remarks

For a long time, multiple deposit expansion involving the money multiplier process was considered a 'monetary law'. In reality, however, it is only a specific phenomenon arising in the early stages of financial system development. Having reviewed the evolution of liquidity management practices in banks over the past century, we arrived at a conclusion that multiple deposit expansion and the money multiplier logic may have existed in past decades but are unlikely to have been operative in the last decades in Russia – and quite likely in developed and monetary sovereign countries too.

In other words, the money multiplier story can be a reality as well as a myth, depending on the institutional evolution of a particular country's banking system and monetary policy framework – which are ultimately interlinked and influencing one another via the global(ized) economy. We restated the classical conditions when the money multiplier exists, and their violation recently explains as well its non-existence nowadays.

In the past, banks were subject to frequent 'runs'; had limited access to central bank facilities due to 'stigma'; could not use the money market to redistribute reserves from those with ample reserves to those who lack them. In addition, reserves tended to shrink with trade balance contraction and vice versa. At the same time, banks could alter the interest rates on their loans and deposits more freely than nowadays. In sum, the money supply may have been exogenous while interest rates were endogenous.

Modern banks do not accumulate reserves in order to lend, and do not strive to maintain zero liquidity gaps (since their main function nowadays is 'liquidity transformation', which is totally incompatible with 'the golden rule' of classical banking). Central banks, in their turn, provide free access to their facilities for financially sound counterparties, covering their liquidity needs in full. Leading central banks try to influence their economies managing the interest rates, not the money supply. As a result, now the money supply is endogenous while interest rates are exogenous.

These conclusions were supported by a number of examples from central banking practice and modern banks' behavior. We revealed that excess reserves in Russia have become negligible and play no significant role in money creation. Neither of multiple deposit expansion criteria (supply-side constraints faced by banks and their inability to circumvent them; absence of demand-side constraints) are applicable recently – to Russia, at least, for the second subperiod under consideration in our empirical work here (2012-2019).

To sum up, our hypothesis is generally in line with Chick's concept of banking sector evolution (Chick, 1992). At the same time, we interpret the obtained results a bit differently: money has always been endogenous; however, an aggressive environment compelled banks to follow very conservative liquidity management strategies in the past. As financial systems developed, they gradually abandoned these inefficient practices. Consequently, money creation and money multipliers are no longer reliable concepts not only for developed countries, but also for a number of emerging markets (including those having underdeveloped financial market and external imbalances), such as Russia. These conclusions support the notion that monetary targeting was not an appropriate regime for Russia in 2000s, while the transition to inflation targeting which started in 2015, was justified.

Our findings are, more generally, consistent with those in Carpenter and Demiralp (2012) based on US data and interpreted against the evolution of the US monetary policy and banking institutional framework. The literature along the lines we pursued, however, has only recently been emerging, to essentially challenge the money multiplier mechanism under the conditions in the wake of the GFC. Much further work, on other countries and theoretically-grounded in microeconomic behavior, remains to be done.

References

[1] **Badarudin, Zatul E., Ahmed M. Khalid, and Mohamed A. Khalid.** 2009. "Money Supply Behavior in Emerging Economies: A Comparative Analysis." *Journal of the Asia Pacific Economy* 14 (4): 331–350.

[2] **Bain, Keith, and Peter Howells.** 2005. *The Economics of Money, Banking and Finance: A European Text*. New York: Palgrave Macmillan.

[3] **Bernanke Ben S., and Alan S. Blinder.** 1992, "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82 (4): 901–921.

[4] **Bindseil, Ulrich.** 2004. *Monetary Policy Implementation: Theory, Past, and Present*. Oxford: Oxford University Press.

[5] **Bindseil, Ulrich.** 2011. *Monetary Policy Operations and the Financial System*. Oxford: Oxford University Press.

[6] **Borio, Claudio.** 2019. "On Money, Debt, Trust and Central Banking." Bank for International Settlements Working Paper No. 763.

[7] **Borio, Claudio, and Piti Disyatat.** 2009. "Unconventional Monetary Policies: An Appraisal." Bank for International Settlements Working Paper No. 292.

[8] **Cagan, Philip.** 1965. *Determinants and Effects of Changes in the Stock of Money, 1875–1960*. New York, London: National Bureau of Economic Research.

[9] **Carpenter, Seth, and Selva Demiralp.** 2012. "Money, Reserves, and the Transmission of Monetary Policy: Does the Money Multiplier Exist?" *Journal of Macroeconomics* 34: 59–75.

[10] **Cetorelli, Nicola, and Linda Goldberg.** 2011. "Global Banks and International Shock Transmission: Evidence from the Crisis." *IMF Economic Review* 59 (1): 41–76.

[11] **Chick, Victoria.** 1992. "The Evolution of the Banking System and the Theory of Saving, Investment and Interest." In *On Money, Method and Keynes: Selected Essays*, edited by Philip Arestis, Sheila C. Dow, and William A. Peniston. Palgrave Macmillan, 193–205 (Chapter 12).

[12] **Crick, W. F.** 1927. "The Genesis of Bank Deposits." *Economica* 20 (Jun.): 191–202.

[13] **Disyatat, Piti.** 2010. "The Bank Lending Channel Revisited." Bank for International Settlements Working Paper No. 297 (February).

[14] **ECB.** 2011. "The Supply of Money – Bank Behaviour and the Implications for Monetary Analysis." European Central Bank *Monthly Bulletin* (October).

[15] **Gedeon, Shirley J.** 1985. "The Post Keynesian Theory of Money: A Summary and an Eastern European Example." *Journal of Post Keynesian Economics* 8 (2): 208–222.

[16] **Goodhart Charles.** 1989. "The Conduct of Monetary Policy." *Economic Journal* 99 (396): 293–346.

[17] **Grishchenko, Vadim O.** 2017. "The Money Multiplier in Modern Financial Systems." *Den'gi I Kredit* 9: 10–17 (in Russian).

[18] **Grishchenko, Vadim O.** 2018. "The Money Multiplier in the Context of Modern Views on Money Creation: Theory and Facts." *Voprosy Economiki* 11: 50–69 (in Russian).

[19] **Humphrey, Thomas M.** 1974. "The Quantity Theory of Money: Its Historical Evolution and Role in Policy Debates." FRB Richmond *Economic Review*, May/June, 2–19.

[20] **Humphrey, Thomas M.** 1987. "The Theory of Multiple Expansion of Deposits: What It Is and Whence It Came." FRB Richmond *Economic Review*, March/April, 3–38.

[21] **Humphrey, Thomas M.** 1989. "Lender of Last Resort: A Concept in History." FRB Richmond *Economic Review*, March/April, 8–16.

[22] **Ireland, Peter N.** 2004. "Money's Role in the Monetary Business Cycle." *Journal of Money, Credit, and Banking* 36 (6): 969–983.

[23] **Jakab, Zoltan, and Michael Kumhof.** 2015. "Banks Are Not Intermediaries of Loanable Funds – and Why This Matters." Bank of England Working Paper No. 529.

[24] **Jordan, Thomas J.** 2018. "How Money is Created by the Central Bank and the Banking System" (speech). Swiss National Bank, Zurich.

[25] **Juurikkala, Tuuli, Alexei Karas, and Laura Solanko.** 2011. "The Role of Banks in Monetary Policy Transmission: Empirical Evidence from Russia." *Review of International Economics* 19 (1): 109–121.

[26] **Lavoie, Marc.** 1984. "The Endogenous Flow of Credit and the Post Keynesian Theory of Money." *Journal of Economic Issues* 18 (3): 771–797.

[27] **Leeper, Eric, and Jennifer Roush.** 2003. "Putting 'M' Back into Monetary Policy." *Journal of Money, Credit, and Banking* 35 (6, Part 2): 1217–1256.

[28] **McKnight, Stephen, and Alexander Mihailov.** 2015. "Do Real Balance Effects Invalidate the Taylor Principle in Closed and Open Economies?" *Economica* 82 (328): 938–975.

[29] **McLeay, Michael, Amar Radia, and Ryland Thomas.** 2014. "Money Creation in the Modern Economy." Bank of England *Quarterly Bulletin*, Q1.

[30] **Meade, James E.** 1934. "The Amount of Money and the Banking System." *Economic Journal* 44 (173): 77–83.

[31] **Mihailov, Alexander.** 2009. "Exchange Rate Pass-through to Prices in Macrodata: A Comparative Sensitivity Analysis." International Journal of Finance & Economics 14 (4): 346–377.

[32] **Minsky, Hyman P.** 1960. *The Pure Theory of Banking*. University of California, Department of Economics (Fall).

[33] **Mishkin, Frederic S.** 2014. *The Economics of Money, Banking, and Financial Markets*, Pearson (Eleventh edition).

[34] **Moore, Basil J.** 1979. "The Endogenous Money Stock." *Journal of Post Keynesian Economics* 2 (1, Fall): 49–70.

[35] **Moore, Basil J.** 1983. "Unpacking the Post Keynesian Black Box: Bank Lending and the Money Supply." *Journal of Post Keynesian Economics* 5 (4): 537–556.

[36] **Moore, Basil J.** .1988. *Horizontalists and verticalists: The Macroeconomics of Credit Money*. Cambridge: Cambridge University Press.

[37] **Moulton H. G.** 1918. "Commercial Banking and Capital Formation (I-IV)." *Journal of Political Economy* 26 (5–7,9).

[38] **Pesaran, Hashem M., and Yongcheol Shin.** 1998. "Generalized Impulse Response Analysis in Linear Multivariate Models." *Economics Letters* 58: 17–29.

[39] **Phillips, Chester Arthur.** 1920. *Bank Credit: A Study of Principles and Factors underlying Advances Made by Banks to Borrowers*. New York: The Macmillan company.

[40] **Ponomarenko Alexey, Elena Vasilieva, and Schobert, Franziska.** 2012. "Feedback to the ECB's Monetary Analysis: The Bank of Russia's Experience with Some Key Tools." European Central Bank Working Paper No. 1471 (September).

[41] **Ponomarenko, Alexey.** 2017. "A Note on Money Creation in Emerging Market Economies." *Journal of Financial Economic Policy* 9 (1): 70–85.

[42] **Ponomarenko, Alexey.** 2019. "Do Sterilized Foreign Exchange Interventions Create Money?" *Journal of Asian Economics* 62: 1–16.

[43] **Prochnow, Herbert V.** 1949. "Bank Liquidity and the New Doctrine of Anticipated Income." *Journal of Finance* 4 (4, Dec): 298–314.

[44] **Samuelson Paul A.** 1948. *Economics*. New York: McGraw-Hill.

[45] **Salmanov Oleg N., Victor M. Zayernyuk, and Olga A. Lopatina.** 2015. "An Analysis of the Impact of Monetary Policy on Bank Lending in Russia." *Asian Social Science* 11 (6).

[46] **Sims, Christopher.** 1980. "Macroeconomics and Reality." *Econometrica* 48: 1–48.

[47] **Vymyatnina, Yulia.** 2006. "How Much Control Does Bank of Russia Have over Money Supply?" *Research in International Business and Finance* 20: 131–144.

[48] **Vymyatnina, Yulia.** 2013. "Money Supply and Monetary Policy in Russia: A Post-Keynesian Approach Revisited." European University in St. Petersburg. Department of Economics. Working Paper Ec-04/13.

[49] **Woodford, Michael.** 2010. "Financial Intermediation and Macroeconomic Analysis." *Journal of Economic Perspectives* 24 (4, Fall): 21–44.

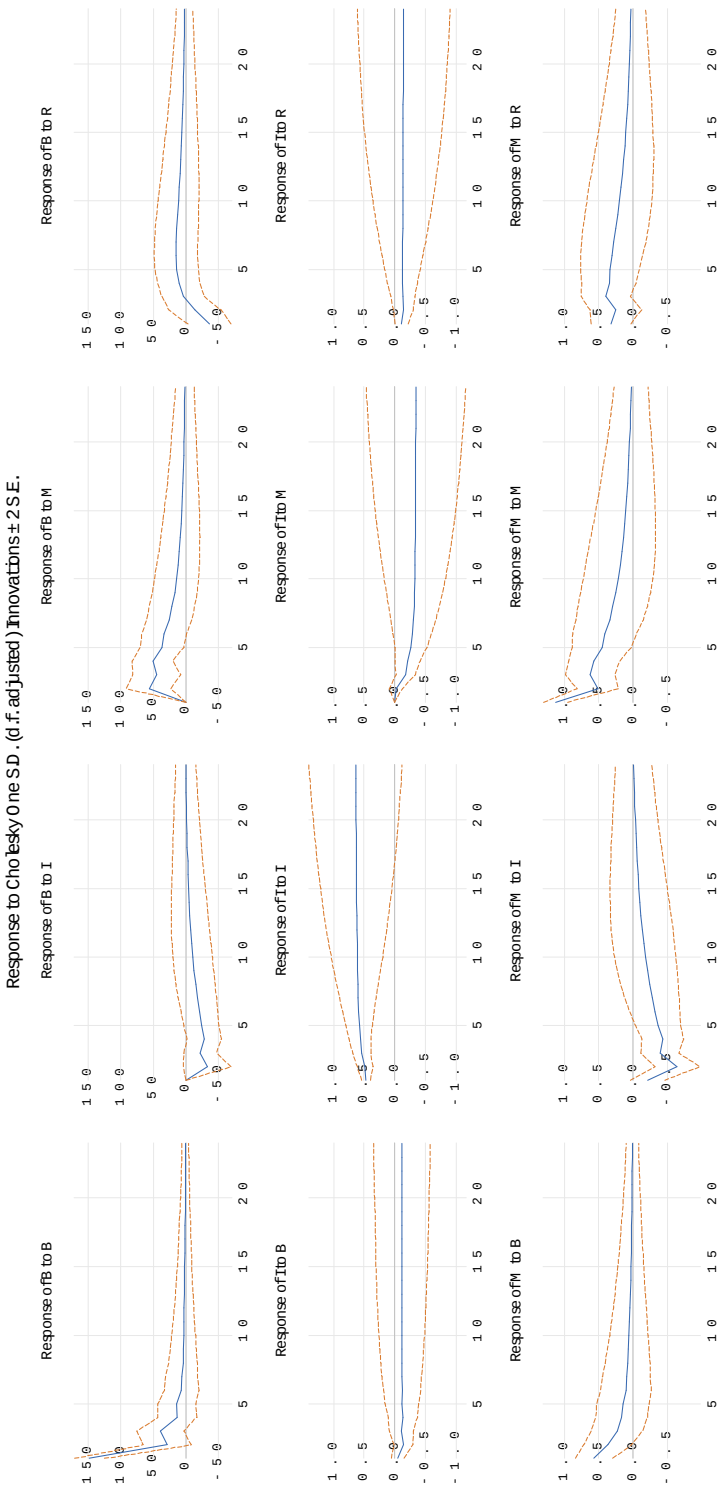


Figure 14: Cholesky Impulse Responses, Main Specification, 1st Subsample (2005:01-2012:02)

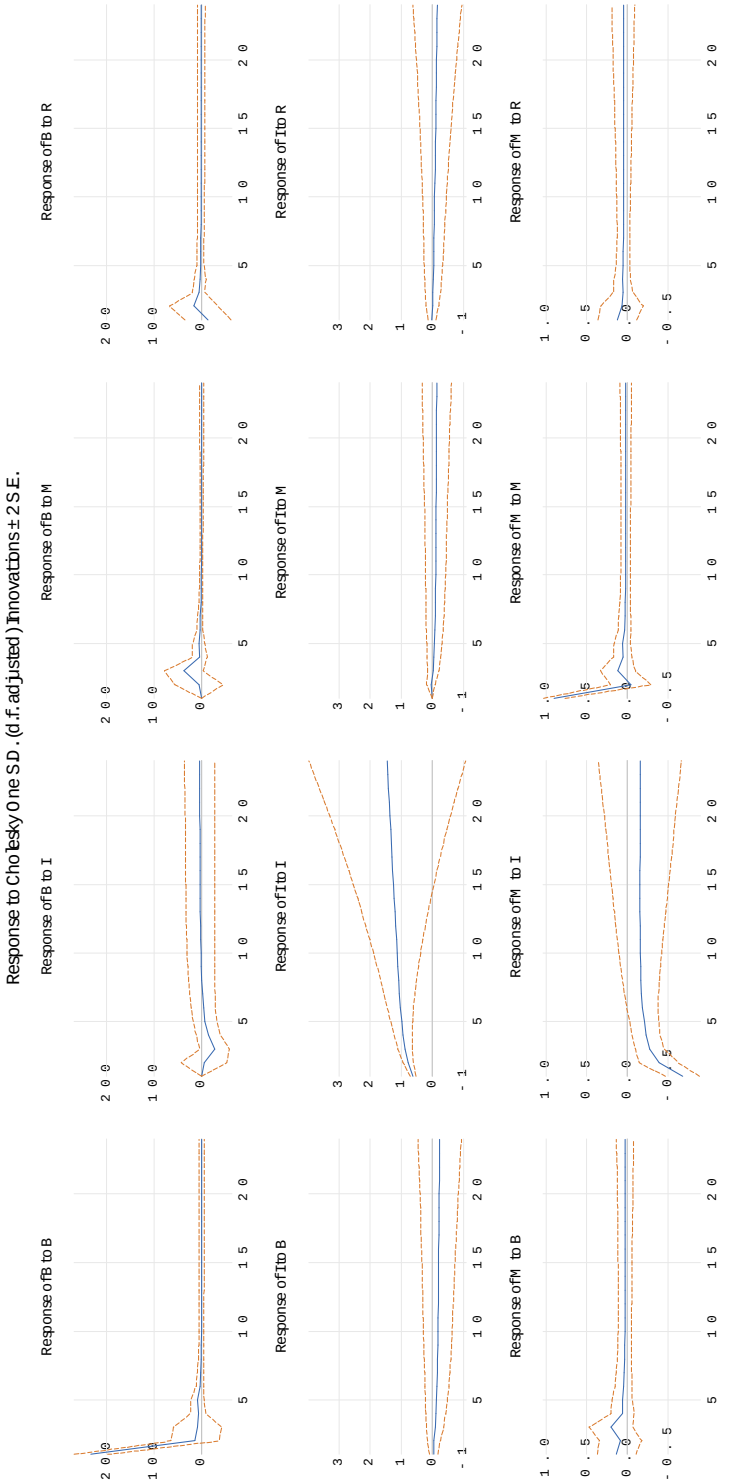


Figure 15: Cholesky Impulse Responses, Main Specification, 2nd Subsample (2012:03-2019:12)

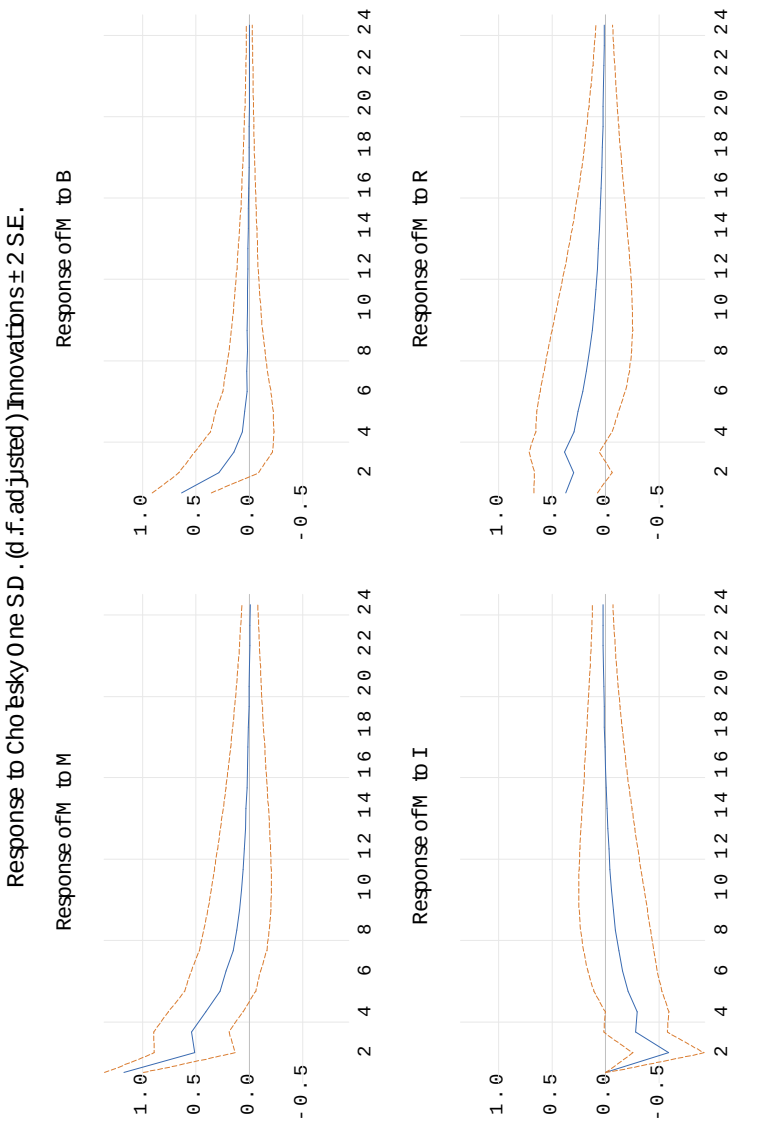


Figure 16: Cholesky Impulse Responses, Alternative Ordering, 1st Subsample (2005:01-2012:02)

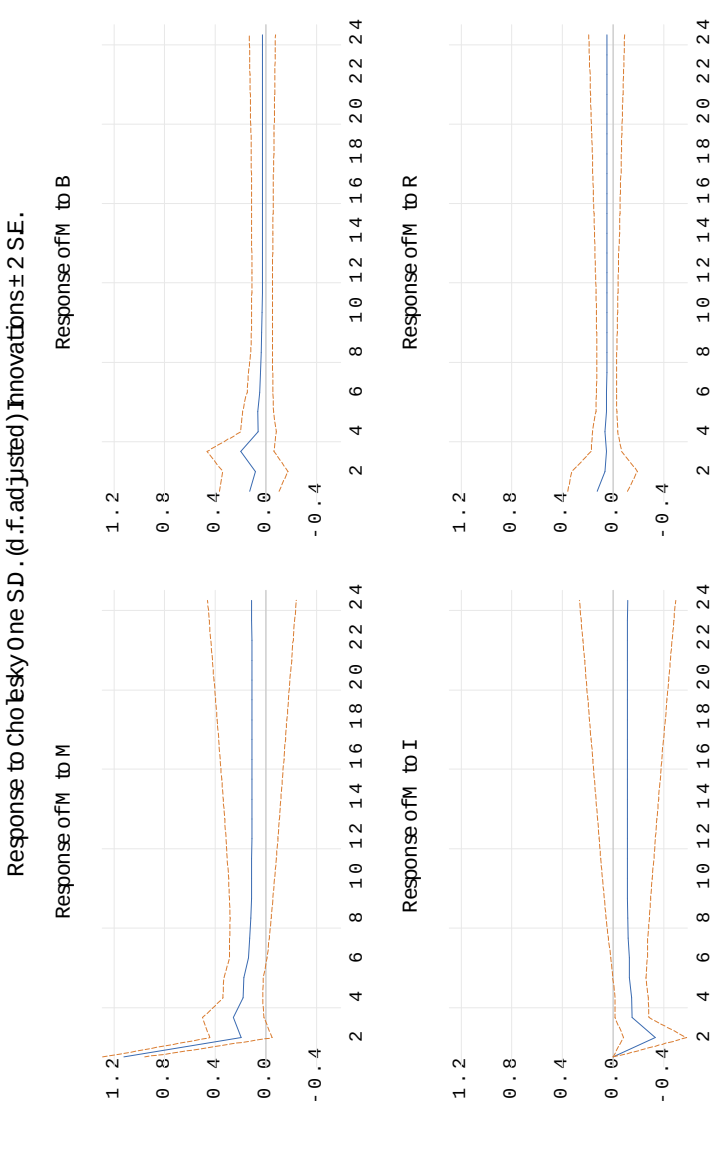


Figure 17: Cholesky Impulse Responses, Alternative Ordering, 2nd Subsample (2012:02-2019:12)

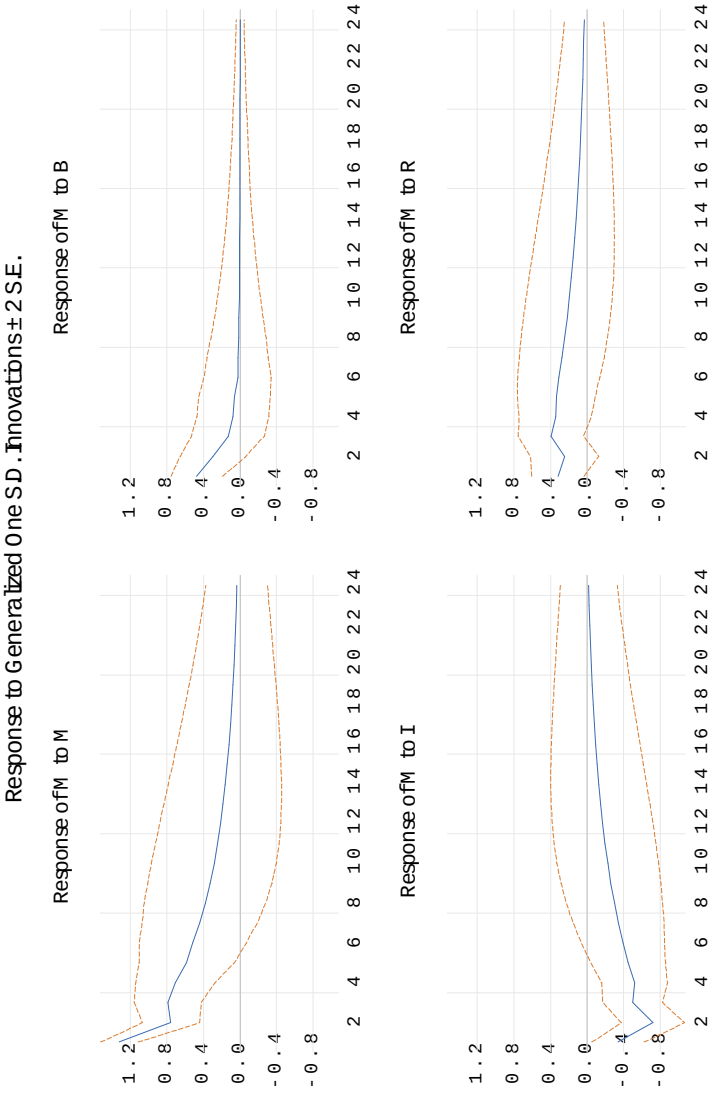


Figure 18: Generalized Impulse Responses, 1st Subsample (2005:01-2012:02)

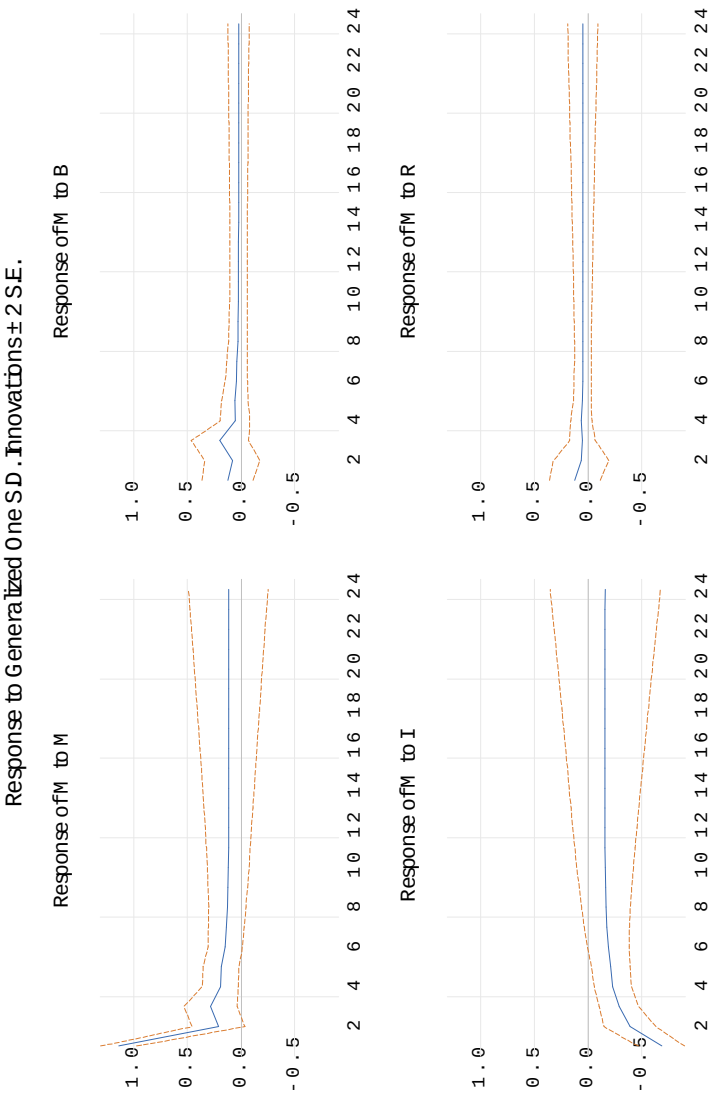


Figure 19: Generalized Impulse Responses, 2nd Subsample (2012:03-2019:12)