UNIVERSITÄT
BERN

Faculty of Business, Economics
and Social Sciences

Department of Economics

Money and Banking with Reserves and CBDC

Dirk Niepelt

22-12

October, 2022

DISCUSSION PAPERS

Schanzeneckstrasse 1
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Money and Banking with Reserves and CBDC*

Dirk Niepelt†

CEPR DP 15457, November 2020
This version: October 28, 2022

Abstract

We analyze retail central bank digital currency (CBDC) in a two-tier monetary system with bank deposit market power and externalities from liquidity transformation. Resource costs of liquidity provision determine the optimal monetary architecture and modified Friedman (1969) rules the optimal monetary policy. Optimal interest rates on reserves and CBDC differ. A calibration for the U.S. suggests a weak case for CBDC in the baseline but a much clearer case when too-big-to-fail banks, tax distortions or instrument restrictions are present. Depending on central bank choices CBDC raises U.S. bank funding costs by up to 1.5 percent of GDP.

JEL codes: E42, E43, E51, E52, G21, G28

Keywords: Central bank digital currency, reserves, two-tier system, bank, liquidity, equivalence

*This paper previously circulated under the title "Monetary Policy with Reserves and CBDC." For valuable comments or discussions, we are grateful to David Andolfatto, Raphael Auer, Marco Bassetto, Javier Bianchi, Florian Böser, Antonio Fatás, Daria Finocchiaro, Martín Gonzalez-Eiras, Timothy Jackson, Arvind Krishnamurthy, Antoine Martin, Cyril Monnet, Thomas Moser, Juan Pablo Nicolini, Galo Nũno, Chris Phelan, Antti Ripatti, Jean-Charles Rochet, Remo Taudien, Harald Uhlig, Stephen Williamson, Cathy Zhang, Shengxing Zhang, as well as conference or seminar participants at the ASSA meetings, Barcelona GSE Summer Forum, Bundesbank, CB&DC Virtual Seminar, CEMFI, CESifo Area Conference on Macro, Money, and International Finance, European Central Bank, Federal Reserve Bank of Minneapolis, Joint Online Macro/Finance Seminar (Academia Sinica, Hong Kong Baptist University, Monash University, National Taiwan University, New York University-Shanghai, Yonsei University, Renmin University of China), National Bank of Ukraine, Philadelphia Workshop on the Economics of Digital Currencies, Riksbank, Swiss National Bank, SNB-CIF Conference, TU Berlin, Universities of Copenhagen, Frankfurt, Konstanz, Liverpool, and Zurich (joint with Federal Institute of Technology), Virtual Finance Theory Seminar, and the 6th International Macroeconomics and Finance Conference (Keio/CEPR). We thank Luca Benati, Han Gao, and Juan Pablo Nicolini for help with the data and Pablo Kurlat for sharing his data.

†University of Bern; CEPR. dirk.niepelt@unibe.ch, www.niepelt.ch.

1 Introduction

The prospect of retail central bank digital currency (CBDC) or "Reserves for All" has revived interest in fundamental questions about money creation and the monetary architecture. In modern economies this architecture exhibits two tiers: Nonbanks use deposit liabilities of banks or claims on deposits as payment instruments, while banks pay each other with reserves that are issued by the central bank. New financial service providers build on these payment rails and exploit synergies but they do not undermine the basic architecture. CBDC, in contrast, subverts the two-tier system by giving the general public direct access to digital central bank money.[1]

Complementary trends propel monetary regime change, with similar effects on banks. After the 2007 financial crisis central banks expanded balance sheets and banks responded to "quantitative easing" and tightened liquidity regulation by backing a larger share of their deposits with reserves. As a consequence money multipliers collapsed—by roughly 50 percent.[2] While "sovereign money" proposals in the tradition of the "Chicago Plan" from the 1930s or the recent Swiss "Vollgeld" initiative to outlaw fractional reserve banking have not succeeded, private retail money creation has lost ground.[3]

This paper focuses on the implications of CBDC for the financial sector and the wider economy; they concern the monetary architecture, liquidity provision, bank funding costs, credit, investment, and seignorage. It analyzes the contemporary two-tier monetary system as well as the likely future, integrated system with deposits, reserves, and CBDC. We address positive questions on the consequences of CBDC as well as normative questions on the optimal monetary system and the optimal policy within such a system.

Our framework builds on Sidrauski (1967), embedding money and banking into the workhorse general equilibrium model. Households value goods, leisure, and the liquidity services provided by deposits and CBDC. Neoclassical firms produce with capital and labor. Banks invest in capital and reserves, issue equity, and exert deposit market power. The central bank issues CBDC and reserves. (We also consider cash and government bonds but they do not play important roles.) Payments require resources and reserve holdings help banks avoid liquidity shortages, which cause bank internal and external (fire-sale) costs. The policy instruments at the government's disposal include the interest rates on reserves and CBDC, or the quantities of these liabilities, as well as deposit subsidies to counteract markup distortions.

Our liquidity centric view of banks emphasizes the creation and management of payment instruments rather than frictions on the lending side. It reflects the focus on mon-

[1] See for example Board of Governors (2022). CBDC re-introduces (rather than newly introduces) noncash public retail payment instruments as central banks offered accounts to nonbanks in the past (BIS, 2018). For an overview of CBDC projects see for example Boar and Wehrli (2021). Granting the general public access to digital central bank money does not require the central bank to interact with retail customers; private payment service providers can play this role (e.g., Auer and Böhme, 2020).

[2] In the U.S., the M3 multiplier declined by 53 percent (and the M1 multiplier by 48 percent) between December 2007 and December 2009 and has not recovered since (source: FRED).

[3] The fall in money multipliers is predominantly a consequence of increased reserve holdings by banks. On the "Chicago Plan," see for example Knight et al. (1933) and Fisher (1935); on the "Vollgeld" initiative, see vollgeld-initiative.ch.

etary architecture, is empirically relevant and analytically innocuous.[4] The resulting equilibrium conditions reduce to the baseline real business cycle model augmented by "pseudo wedges," which summarize the effects of policy instruments and payment system characteristics. In the limiting case without liquidity demand the pseudo wedges disappear.

Our positive analysis starts with a comparison of the transmission mechanisms in the single- and two-tier systems. In the conventional, two-tier system the mechanism operates through the cost structure of banks and their price setting, which depend on subsidies, the spread on reserves, structural factors that determine the benefits for banks of holding reserves, as well as bank market power. To induce banks to raise the equilibrium deposit rate the central bank may increase interest on reserves or deposit subsidies; the former is fiscally cheaper due to the externalities from reserve holdings. An increase in the liquidity premium on reserves triggers quantitatively important substitution, income, and Pareto substitutability effects, divorcing investment and consumption (and output) growth.

In a single-tier, CBDC based system monetary transmission is more direct. Rather than providing incentives for banks the central bank directly controls the spread and all pseudo wedges. Finally, in a system with both deposits and CBDC the policy mix is critical. A CBDC quantity target affects the elasticity of deposit funding and modifies banks' price setting. A target for the composition of real balances affects bank profits but not their price setting. And a target for the CBDC spread forces banks to follow suit or drives them out of business.

We extend the Brunnermeier and Niepelt (2019) equivalence result to incorporate resource costs of payments. We show that, when the public and the private sector provide liquidity equally efficiently, portfolio shifts out of deposits into CBDC do not undermine bank lending nor affect the allocation as long as the central bank passes its new funding back to banks at an equivalent rate. We emphasize the generality of this result which extends far beyond our workhorse framework. But we also use the result to identify potential sources of nonequivalence and to categorize the nascent literature. One potential source concerns the fact that central bank loans under the equivalent policy do not require collateral. We argue that, rather than rendering the equivalence result "unrealistic," this points to a potential inconsistency of central bank policies.

Under the equivalent policy banks do not supply liquidity any longer. Nevertheless, the central bank funds them at the equivalent loan rate (which reflects the deposit rate and other factors) rather than an "illiquid" market rate. This triggers questions about the political viability of the equivalent policy. We characterize and quantify the "funding-cost-reduction-at-risk" for U.S. banks, i.e., the difference between bank funding costs at the illiquid risk-free rate and the equivalent loan rate. We find that since 1999, the funding-cost-reduction-at-risk has varied substantially and at times exceeded 1.5 percent of GDP, suggesting that banks face major political risks from the introduction of CBDC.

Next, we turn to the normative analysis. We start by characterizing the allocation chosen by a social planner that is constrained by the production and payment technologies.

[4] Not only do banks hold securities in addition to loans but a large share of loans are mortgages which can be securitized (Piazzesi and Schneider, 2021). Introducing lending frictions would not change the key conclusions of the analysis unless such frictions directly interacted with bank liabilities; see section 5.

The planner provides payment instruments up to the point where the marginal liquidity benefit equals the marginal social costs, that is, it follows the Friedman (1969) rule which commonly abstracts from resource costs of managing payments. As for the source of liquidity, the planner relies on the monetary architecture that generates the lowest effective resource costs.

We show that the Ramsey government implements the planner allocation. In a CBDC-based system this is simple: The central bank sets the liquidity premium on CBDC at a level that reflects social costs and it prices banks out of the market. In a two-tier system the situation is more challenging since the government needs to correct two distortions—due to market power and externalities in the banking sector—but this is feasible given its two instruments. The optimal liquidity premium on reserves induces banks to choose the efficient quantity of reserves even in the presence of externalities; and the optimal deposit subsidy induces banks to charge the efficient liquidity premium on deposits.

Importantly, bank market power does not necessarily imply a positive optimal deposit subsidy, for the latter reflects both frictions in the banking sector: Stronger market power requires a higher subsidy as banks must be encouraged to lengthen their balance sheets, but a stronger reserves externality demands the opposite. Intuitively, when reserves generate external benefits the Ramsey policy increases the interest rate on reserves, and to sterilize the effect on the deposit margin it lowers the deposit subsidy.

Under functional form assumptions we solve for the Ramsey policy rules. When we also impose our preferred calibration we find that reserves should pay nearly the risk-free interest rate; deposits should (still) be subsidized; and deposits should pay roughly seventy basis points less than the risk-free rate. Operating a two- rather than single-tier payment system requires slightly fewer resources; the cost advantage of the two-tier system equals roughly seven percent. The calibrated baseline model thus provides no rationale for a CBDC.

This changes when we introduce additional features such as admissibility constraints, information and commitment problems in the face of too-big-to-fail banks, or tax distortions. To understand the role of admissibility constraints, we consider a scenario in which deposits are the resource-efficient retail means of payment but a deposit subsidy/tax is not admissible. Optimal liquidity provision then requires an alternative instrument to complement interest on reserves. While a circulating CBDC is not an option (because by assumption it would waste resources) a suitably chosen CBDC interest rate target is, because it can induce banks to adjust the deposit rate even if CBDC does not circulate (Andolfatto, 2021).

We show that this mechanism only operates when the unconstrained Ramsey policy subsidizes deposits. When the unconstrained policy taxes deposits (because strong externalities require a high interest rate on reserves), a CBDC interest rate target cannot replace the missing deposit subsidy/tax instrument because it cannot induce banks to *lower* the deposit rate.[5] In such a situation, CBDC only provides second-best alternatives. We investigate a range of them, including a CBDC quantity target and a target for the composition of real balances.

[5]In Andolfatto (2021) the mechanism operates without qualification because no externalities are present.

We also emphasize information and commitment problems in the face of too-big-to-fail banks. We argue that such frictions can rationalize a circulating CBDC even when an ideal two-tier system has lower resource costs than a single-tier system and deposit subsidies/taxes are admissible. Intuitively, when liquidity transformation by too-big-to-fail banks is sufficiently distorted it does not suffice to push banks to adjust their price setting. Constrained efficiency rather requires CBDC to crowd out bank liquidity provision.

To establish that deadweight losses of taxation or regulation provide another rationale for CBDC we focus on the government's budget. We show that unlike in a single-tier system, the optimal monetary policy in a two-tier system requires fiscal resources (or regulation) to correct market failures. Ceteris paribus, a single-tier payment system thus causes fewer deadweight losses than a two-tier system—even outside crisis periods. This can turn the case for a two-tier system on the grounds of lower resource requirements on its head once deadweight losses are also accounted for. In fact, tax and similar distortions may not only rationalize a *circulating* CBDC: Independently of resource costs, they rationalize a *noncirculating* CBDC that disciplines banks without generating deadweight burdens.

A recurrent theme in our first- and second-best analyses is that the two central bank liabilities, reserves and CBDC, should pay different interest rates—independently of whether CBDC actually circulates or only serves to discipline banks. Intuitively, the spread on a circulating means of payment should reflect payment operations costs and externalities, and these generally differ between reserves and CBDC. And when CBDC does not circulate but serves to discipline banks then its interest rate should reflect the central bank's target for the deposit rate which reflects factors beyond the social costs of reserves. Implementing the optimal policy therefore requires the government to price discriminate between wholesale and retail users of central bank liabilities.

Finally, we review several other arguments brought up in CBDC discussions, concerning price rigidity, the effective lower bound, stimulus payments, bank lending frictions, intermediary asset pricing, and monetary policy targeting. We argue that while these factors would have additional implications they would not substantially alter the conclusions drawn from our liquidity centric analysis.

Related Literature The paper relates to the literature on money multipliers (Phillips, 1920), two-tier monetary systems, and inside (bank issued) vs. outside (government supplied) money (Gurley and Shaw, 1960). Tobin (1963; 1969; 1985) discusses the fractional reserve banking system and proposes a precursor to CBDC. Benes and Kumhof (2012) find in a DSGE model that fractional reserve banking raises instability and debt levels. Chari and Phelan (2014) emphasize negative externalities of fractional reserve banking when central bank money is scarce while Taudien (2020) argues that inside money fosters production by lowering producers' financing costs. Faure and Gersbach (2018) compare allocations with and without private money creation. Jackson and Pennacchi (2021) contrast liquidity (safe asset) creation by the private and the public sector.

Building on literatures in macroeconomics and banking we assume that reserves affect banks' operating costs (e.g., Bolton et al., 2020; Ozdenoren et al., 2021; Van den Heuvel, 2019; Vandeweyer, 2019).[6] In Kiyotaki and Moore (2019) a liquid security, like reserves, relaxes resalability constraints and reduces costs. Bianchi and Bigio (2020) model the portfolio choice of banks and monetary policy transmission in a frictional interbank market and Parlour et al. (2022) analyze the "liquidity externality" that arises because banks issue means of payment that may be redeemed at other banks. Our micro foundation for the role of reserves emphasizes fire sales (Shleifer and Vishny, 1992; Stein, 2012) and implies positive externalities.

Our equivalence result builds on Brunnermeier and Niepelt (2019) and seminal earlier contributions (Modigliani and Miller, 1958; Barro, 1974; Wallace, 1981; Bryant, 1983; Chamley and Polemarchakis, 1984; Sargent, 1987). Relative to this literature we emphasize liquidity considerations and resource costs, focus on the CBDC application, and quantify the bank funding cost reduction afforded by deposit creation. The equivalence result also identifies potential sources of nonequivalence and thus, serves to categorize the nascent general equilibrium literature on CBDC.

Williamson (2019) and Böser and Gersbach (2020) analyze the role of central bank collateral requirements and collateral scarcity. In Piazzesi and Schneider (2021) central bank balance sheet length is costly and only banks can offer contingent, on-demand liquidity. In Keister and Sanches (2022) the central bank injects CBDC by transfer rather than absorbing deposits in exchange, promoting exchange but crowding out deposits and investment.[7] In the DSGE models studied by Kumhof and Noone (2021) and Barrdear and Kumhof (2022) the central bank issues CBDC in exchange for government bonds, preventing the deposit-CBDC substitution characterized in the equivalence result.[8] Other sources of nonequivalence relate to information (Keister and Monnet, 2020; Niepelt, 2020b), changes in the asset span (Brunnermeier and Niepelt, 2019) and market participation (Benigno et al., 2022; Ferrari Minesso et al., 2022), or politics (Niepelt, 2021). Schilling et al. (2020) analyze conflicts between allocative efficiency, price stability, and financial stability that are present in monetary economies including a CBDC based system. Design options for CBDC are the focus of Kahn et al. (2018), Bindseil (2020), and Auer and Böhme (2020), among others.

Following Klein (1971), Monti (1972) and a more recent literature (e.g., Drechsler et al., 2017) we stipulate a noncompetitive deposit market. In Andolfatto (2021) the introduction of CBDC leads noncompetitive banks to raise the deposit rate, with positive effects on financial inclusion, and in Garratt and Zhu (2021) this affects the market structure.[9] Chiu et al. (2019) quantitatively assess the implications of CBDC in a framework that combines elements of Andolfatto (2021) and Keister and Sanches (2022); they find nonlinear effects of the CBDC interest rate on bank lending rates. We show that the disciplining role of CBDC must be qualified: CBDC may help redress deposit rates when they are inefficiently low but not when they are inefficiently high and reserves generate externalities. We also show that deadweight losses of taxation or regulation make a

[6]See also Cúrdia and Woodford (2009).

[7]In Keister and Sanches (2022) the central bank issues both a "cash-" and a "deposit-like" CBDC.

[8]See Burlon et al. (2022) for another quantitative DSGE analysis.

[9]Garratt and van Oordt (2021) and Garratt and Lee (2021) analyze the disciplining effect of CBDC on agents that exploit consumer information (see also Kahn et al., 2005).

disciplining, non-circulating CBDC preferable to corrective subsidies or regulation.

Our analysis complements large literatures building on "New Keynesian" and "New Monetarist" frameworks (Woodford, 2003; Galí, 2015; Rocheteau and Nosal, 2017). Unlike the former we emphasize the role of money as means of payment (Tobin, 1969) and we abstract from nominal rigidities (but see section 5); compared to the latter and in line with the literature following Sidrauski (1967) (e.g., Di Tella, 2020) we emphasize general equilibrium implications relative to the market micro structure underlying the supply and demand for liquidity.

Structure of the Paper Section 2 lays out the monetary economy. In section 3 we characterize equilibrium, analyze monetary policy transmission, establish the equivalence result, and compute the funding-cost-reduction-at-risk. Section 4 contains the normative analysis of the baseline model and its extensions. Section 5 briefly discusses further extensions and section 6 concludes.

2 Model

We consider a production economy with a continuum of mass one of homogeneous infinitely-lived households that own (and are served by) banks and competitive firms. Monetary and fiscal policy is set by a consolidated government/central bank.

The environment differs threefold from the monetary economy in Sidrauski (1967). First, we introduce banks. Second, we introduce multiple means of payment. At the retail level there are deposits, which are issued by banks, and CBDC or "reserves for all," which are issued by the central bank. At the wholesale level there are reserves, which are also issued by the central bank but exclusively serve as means of payment for banks. (We discuss the role of cash below.) Finally, we introduce costs of operating the payment system[10] as well as of engaging in liquidity transformation. The three extensions relative to Sidrauski (1967) are to represent key features of the monetary architecture in modern economies in which banks settle the payments of their customers with reserves through a central-bank-run clearing system.

The analysis does not impose restrictions on the sources of aggregate risk;[11] all parameters or functions indexed by time may also depend on histories. To keep the notation simple we only let a few parameters and functions explicitly depend on time and histories.

[10]These costs can alternatively be interpreted as costs of managing assets backing the payment instruments.

[11]Di Tella (2020) analyzes idiosyncratic risk in a Sidrauski (1967) type framework. He finds that idiosyncratic risk shocks and risk aversion shocks have similar effects.

2.1 Households

The representative household takes prices, returns, profits, and taxes as given and solves

$$\max_{\{c_t,x_t,k_{t+1},m_{t+1},n_{t+1}\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}_0[u(c_t, z_{t+1}, x_t)]$$

$$\text{s.t.} \quad k_{t+1} + m_{t+1} + n_{t+1} = k_t R_t^k + m_t R_t^m + n_t R_t^n + w_t(1 - x_t) + \pi_t - c_t - \tau_t, \quad (1)$$

$$k_{t+1}, m_{t+1}, n_{t+1} \geq 0.$$

Here, c_t and x_t denote household consumption of the good and leisure at date t, respectively, and z_{t+1} denotes "effective real balances" carried from t into $t+1$.

Effective real balances are a weighted sum of CBDC or "money," m_{t+1}, and deposits, n_{t+1},

$$z_{t+1} \equiv \lambda_t m_{t+1} + n_{t+1},$$

where the parameter $\lambda_t > 0$ indexes the liquidity benefits of money relative to deposits, reflecting many potential factors such as privacy protection or convenience of use. Our assumption that money and deposits are close substitutes is natural and consistent with the data. On the one hand, Nagel (2016), among others, finds a very high elasticity of substitution between deposits and other liquid assets. On the other hand, the majority of central banks considering the introduction of CBDC anticipate an account based technology; CBDC accounts would have to meet the same anti-money-laundering standards as deposit accounts; and the user experience would be largely the same for CBDC or deposit based digital payments. Against this background, we view $\lambda_t = 1$ as a plausible benchmark assumption but we nevertheless allow for $\lambda_t \neq 1$ and for variation of λ_t across time or histories. Some results even hold when λ_t is a function of m_{t+1} or n_{t+1} as we will point out.

We abstract from household cash holdings. Except for effective-lower-bound considerations, which are secondary in our model without price rigidities, including cash as a third retail means of payment would be largely irrelevant. We discuss this in more detail at the end of the section.

The felicity function u is increasing, strictly concave, and satisfies Inada conditions. The discount factor β lies strictly between zero and unity. As is well known, this "money in the utility function" specification represents a broad set of monetary frictions and flexibly generates a demand for liquidity.[12] All key results of the paper are robust to changing the source of money demand and thus, the reason why the Modigliani and Miller (1958) theorem does not apply for issuers of real balances. What matters is not *why* households value liquidity but rather *that* they do.

Equation (1) represents the household budget constraint. The household invests in capital, k_{t+1}, as well as real balances; pays for consumption and taxes, τ_t; and funds these

[12] One such friction is a "shopping time" friction which renders real balances helpful to economize on time spent shopping; see Saving (1971), Feenstra (1986), McCallum and Goodfriend (1987), and Croushore (1993). Recent papers motivating "money in the utility function" specifications include Van den Heuvel (2008), Krishnamurthy and Vissing-Jorgensen (2012), Stein (2012), Van den Heuvel (2019), Begenau (2020), Di Tella (2020) or Jackson and Pennacchi (2021).

outlays with wage income, distributed profits, π_t, and the gross return on its portfolio. Wage income equals the product of the wage, w_t, and labor supply, $1-x_t$. The portfolio return consists of the returns on capital, $k_t R^k_t$, money, $m_t R^m_t$, and deposits, $n_t R^n_t$.

The gross rates of return on money and deposits, R^m_t and R^n_t respectively, reflect both nominal interest rates and inflation. To keep the notation for spreads simple (see below) we assume that R^m_t and R^n_t are risk-free—i.e., inflation risk is negligible—but we relax this assumption when proving equivalence. The gross rate of return on capital, R^k_t, may be risky.[13]

In equilibrium capital holdings and real balances are strictly positive. The household's optimality conditions for k_{t+1}, m_{t+1}, n_{t+1}, and x_t are given by

$$\begin{aligned} u_c(c_t, z_{t+1}, x_t) &= \beta \mathbb{E}_t[R^k_{t+1} u_c(c_{t+1}, z_{t+2}, x_{t+1})], && (2)\\ u_c(c_t, z_{t+1}, x_t) &\geq \beta R^m_{t+1} \mathbb{E}_t[u_c(c_{t+1}, z_{t+2}, x_{t+1})] + \lambda_t u_z(c_t, z_{t+1}, x_t), \quad m_{t+1} \geq 0, \\ u_c(c_t, z_{t+1}, x_t) &\geq \beta R^n_{t+1} \mathbb{E}_t[u_c(c_{t+1}, z_{t+2}, x_{t+1})] + u_z(c_t, z_{t+1}, x_t), \quad n_{t+1} \geq 0, \\ u_x(c_t, z_{t+1}, x_t) &= u_c(c_t, z_{t+1}, x_t) w_t, && (3) \end{aligned}$$

respectively. The weak inequality in the Euler equation for m_{t+1} or n_{t+1}, respectively, holds with equality if m_{t+1} or n_{t+1} is strictly positive.

To express the Euler equations for m_{t+1} and n_{t+1} more compactly define the risk-free interest rate, R^f_{t+1}, as

$$R^f_{t+1} \equiv 1/\mathbb{E}_t[\mathrm{sdf}_{t+1}],$$

where $\mathrm{sdf}_{t+1} \equiv \beta u_c(c_{t+1}, z_{t+2}, x_{t+1})/u_c(c_t, z_{t+1}, x_t)$ denotes the stochastic discount factor. When the household holds payment instruments of type $i \in \{m, n\}$ then the associated first-order condition reads

$$\lambda^i_t u_z(c_t, z_{t+1}, x_t) = u_c(c_t, z_{t+1}, x_t) \left(1 - \frac{R^i_{t+1}}{R^f_{t+1}}\right), \tag{4}$$

where $\lambda^m_t \equiv \lambda_t$ and $\lambda^n_t \equiv 1$. When the household holds both payment instruments then

$$R^f_{t+1} - R^m_{t+1} = \lambda_t (R^f_{t+1} - R^n_{t+1}). \tag{5}$$

According to equation (4) payment instrument i enjoys a liquidity premium when $\lambda^i_t u_z(c_t, z_{t+1}, x_t) > 0$. We denote this liquidity premium on payment instrument i by

$$\chi^i_{t+1} \equiv 1 - \frac{R^i_{t+1}}{R^f_{t+1}}.$$

Equivalently, $-\chi^i_{t+1}$ equals the spread on payment instrument i compared with a risk-free bond that does not provide liquidity services. When the household holds both payment instruments then, according to equation (5), the liquidity premium on money exceeds the premium on deposits if money is more liquid than deposits ($\lambda_t > 1$), and vice versa.

[13]Variables c_t, x_t, k_{t+1}, m_{t+1}, n_{t+1}, z_{t+1}, R^k_t, R^m_{t+1}, R^n_{t+1}, w_t, π_t, τ_t and parameter λ_t are measurable with respect to information available at date t.

2.2 Banks

Banks issue deposits and equity to fund investments in capital and reserves. We abstract from government bonds in bank balance sheets. Including bonds as a third asset category would be largely irrelevant for the analysis as we discuss at the end of the section.

A household may only hold deposits with the single bank in its home region of which there exist finitely many of equal size. As a consequence banks are monopsonists in deposit markets. Regional borders do not restrain any other type of transaction and households are residual claimants to aggregate bank profits. We introduce bank market power in deposit markets because it is empirically relevant (e.g., Drechsler et al., 2017) and frequently cited as a motivation to introduce CBDC. Its main implication is that banks reduce deposit rates to extract rents. Households accept this markdown (up to a point) because they value the liquidity services of deposits.

The exact form of market power is secondary for this mechanism; we adopt the monopsony assumption for convenience. Alternatively, we could assume that the banking sector is monopsonistically competitive such that households hold a composite of deposits (see Ulate, 2021). Or we could posit that several banks in a region compete à la Cournot, giving rise to optimality conditions that are nearly identical to the conditions derived below.[14] In either case our central results would change minimally, for two reasons. First, it is the elasticity of deposit funding, not the preference or market structure underlying this elasticity, which is important for the monetary transmission mechanism. And second, our analysis concerns the substitutability between deposits and money, not the substitutability between different types of deposits.

On the asset side of their balance sheets banks are price takers. We make this assumption because it helps focus on the key questions of interest which concern bank funding and liquidity, and because it is not critical. While market power in lending markets would give rise to markups on lending rates in addition to the model implied markdowns on deposit rates (Klein, 1971; Monti, 1972) the interaction between the two frictions would be unimportant for our results; moreover, our assumption that competitive households also invest in capital leaves no room for bank loan market power (see section 5).

Per unit of deposit funding banks require resources ν to make payments on behalf of their deposit customers.[15] To introduce a role for reserves, r_{t+1}, we assume that liquidity transformation requires bank resources as well; larger reserve holdings relative to deposits reduce the extent of liquidity transformation and thus, these costs. (We do not stipulate a minimum reserves requirement.[16]) Our liquidity centric narrative of the benefits of reserve holdings is consistent with micro foundations in Van den Heuvel (2019), Bianchi

[14] See Freixas and Rochet (2008). Cournot competition scales the elasticity of deposit funding perceived by an individual bank in the same way as a change in preferences.

[15] Operating costs could depend on the number of transactions in addition to balance sheet size, i.e., ν could depend on velocity and thus, on interest rates. We abstract from this effect as it would not alter the key arguments.

[16] A minimum reserves requirement could act similarly as the costs of liquidity substitution that reserves help to avoid, which we consider. An optimum minimum reserves requirement could place banks individually in a corner while being optimal for the banking sector as a whole, due to externalities.

and Bigio (2020), Ozdenoren et al. (2021), and other work.[17] The specific micro foundation we develop in appendix A builds on a fire sale narrative along the lines of Shleifer and Vishny (1992) and Stein (2012): When, as a consequence of liquidity transformation, a bank lacks reserves to settle payments with other financial institutions it sells capital to the central bank, which acts as lender of last resort.[18] But asymmetric information or related frictions depress the price of capital when other institutions find themselves in the same situation and fire sell. This has two consequences: reserve holdings generate a positive pecuniary externality. And in the face of fire sale risk each bank engages in costly precautionary liquidity substitution, which is more expensive when the bank itself or its peers hold fewer reserves.

Formally, a bank's resource costs of liquidity substitution equal $\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1})$ per deposit, and they are strictly decreasing in the bank's reserves-to-deposits ratio (the "liquidity ratio"), $\zeta_{t+1} \equiv r_{t+1}/n_{t+1}$, as well as the aggregate ratio among banks, $\bar{\zeta}_{t+1} \equiv \bar{r}_{t+1}/\bar{n}_{t+1}$. When both the bank and its peers do not engage in liquidity transformation but only invest in reserves—i.e., when banks are "narrow banks" and deposits amount to "synthetic CBDC"—then liquidity substitution is not needed, $\omega_t(1, 1) = 0$. For technical reasons, we assume that in equilibrium $\omega_{11,t} + \omega_{12,t} > 0$ and $\omega_{21,t} + \omega_{22,t} > 0$.[19] Appendix A derives our preferred specification for ω_t in light of the micro foundations sketched above.

Each regional bank is long-lived but its program consists of a sequence of static problems, which involve a choice of deposits and reserve holdings that also determines capital investment and equity issuance. Formally, the date-t program of a bank reads[20]

$$
\begin{aligned}
\max_{n_{t+1}, r_{t+1}} \quad & \pi^b_{1,t} + \mathbb{E}_t[\mathrm{sdf}_{t+1}\ \pi^b_{2,t+1}] \\
\text{s.t.} \quad & \pi^b_{1,t} = -n_{t+1}(\nu + \omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1}) - \theta_t), && (6) \\
& \pi^b_{2,t+1} = (n_{t+1} - r_{t+1})R^k_{t+1} + r_{t+1}R^r_{t+1} - n_{t+1}R^n_{t+1}, && (7) \\
& R^n_{t+1} \text{ reflects deposit funding schedule,} \\
& n_{t+1} \geq 0,
\end{aligned}
$$

where $\pi^b_{1,t}$ and $\pi^b_{2,t+1}$ denote the cash flows at dates t and $t+1$, and θ_t denotes a deposit subsidy. The first constraint relates the cash flow in the first period to the payment operations and liquidity substitution costs net of subsidies; this negative cash flow must be financed by equity issuance. The second constraint relates the cash flow in the second period, which equals the return on equity, to the gross yield on the bank's portfolio; that is, on the return on capital investment, $n_{t+1} - r_{t+1}$, reserve holdings, and deposits. The

[17] It is also consistent with narratives that emphasize the costs of managing bank assets. See for example Cúrdia and Woodford (2009) and the work quoted there.

[18] In the presence of heterogeneous liquidity shocks banks might also sell capital on interbank markets.

[19] We denote the partial derivatives of ω_t with respect to the first and second argument, respectively, by $\omega_{1,t}$ and $\omega_{2,t}$ and we use similar notation for higher order derivatives. The two inequalities given in the text are satisfied, for example, when ω_t is strictly convex and $\omega_{12,t} = 0$ or $\omega_{11,t} \geq \omega_{22,t}$. We focus on symmetric bank choices and consider parameters for which banks choose interior levels of ζ_{t+1}.

[20] Variables $r_{t+1}, \zeta_{t+1}, \mathrm{sdf}_t, R^r_{t+1}, \pi^b_{1,t}, \pi^b_{2,t}, \theta_t$, and ω_t are measurable with respect to information available at date t. We do not normalize the portfolio positions by the number of banks; that is, we state the conditions as they apply for the banking sector as a whole.

third constraint reflects the fact that the bank is a monopsonist; it takes the deposit funding schedule rather than the deposit rate as given. We assume, and later verify, that the funding schedule is differentiable.

Note that the gross rate of return on bank equity, which equals $-\pi^b_{2,t+1}/\pi^b_{1,t}$ and exceeds the required rate of return, is increasing in the liquidity premium on deposits and decreasing in the liquidity premium on reserves. If investors were to inject more equity than $-\pi^b_{1,t}$ then the excess funds would optimally be invested in capital. In other words, inframarginal bank equity funds the monopsonist's portfolio while marginal bank equity funds capital investment. Note also that the bank's program yields a determinate equilibrium leverage ratio as the Modigliani and Miller (1958) theorem does not apply: Deposit relative to equity funding equals $-n_{t+1}/\pi^b_{1,t} = 1/(\nu + \omega_t(\cdot) - \theta_t)$. While banks wish to issue some equity in order to being able to engage in profitable deposit business, they do not have incentives to issue "too much" equity.

The marginal effect of n_{t+1} on the bank's objective is given by

$$-(\nu + \omega_t(\cdot) - \theta_t) + \omega_{1,t}(\cdot)\zeta_{t+1} + \mathbb{E}_t[\text{sdf}_{t+1}(R^k_{t+1} - R^n_{t+1} - n_{t+1}R^{n}_{t+1}{}'(n_{t+1}))].$$

Using the household's Euler equation this implies that an active bank ($n_{t+1} > 0$) satisfies

$$-(\nu + \omega_t(\cdot) - \theta_t) + \omega_{1,t}(\cdot)\zeta_{t+1} + \chi^n_{t+1} = n_{t+1}R^{n}_{t+1}{}'(n_{t+1})/R^f_{t+1}. \tag{8}$$

The left-hand side of equality (8) represents the marginal profit from deposit issuance, holding the interest rate on deposits constant: A marginal unit of deposits generates net operating and liquidity substitution costs $\nu + \omega_t(\cdot) - \theta_t$ and it increases the liquidity substitution costs for inframarginal units, but it also yields a gain if the deposit liquidity premium is positive, $\chi^n_{t+1} > 0$. The right-hand side of the equality equals the profit loss that results because higher deposit issuance forces the bank to increase R^n_{t+1}. Equation (8) simplifies to

$$\chi^n_{t+1} - (\nu + \omega_t(\cdot) - \theta_t - \omega_{1,t}(\cdot)\zeta_{t+1}) = \frac{1}{\eta_{n,t+1}}\frac{R^n_{t+1}}{R^f_{t+1}},$$

where $\eta_{n,t+1}$ denotes the elasticity of deposit funding with respect to R^n_{t+1} (see Klein, 1971; Monti, 1972). This elasticity may depend on central bank choices, in particular on whether—and how elastically—the central bank supplies m_{t+1}. We address this in detail in subsequent sections.

An active bank holds reserves. The corresponding first-order condition reads

$$-\omega_{1,t}(\zeta_{t+1}, \bar{\zeta}_{t+1}) = 1 - \frac{R^r_{t+1}}{R^f_{t+1}}. \tag{9}$$

Intuitively, the optimal choice of reserves equalizes the (private) gain from lower liquidity substitution costs and the return loss due to the liquidity premium on reserves, $\chi^r_{t+1} \equiv 1 - R^r_{t+1}/R^f_{t+1}$. Since in equilibrium $\zeta_{t+1} = \bar{\zeta}_{t+1}$ equation (9) implies a unique mapping from the opportunity cost of holding reserves to the equilibrium reserves-to-deposits ratio, which we write as $\zeta_{t+1} = \omega^{-1}_{1,t}(-\chi^r_{t+1})$.[21] Naturally, our framework therefore implies that

[21] By the mean value theorem, for any $\varepsilon > 0$ there exists a $\iota \in (0, \varepsilon)$ such that $\omega_{1,t}(\zeta + \varepsilon, \zeta + \varepsilon) = \omega_{1,t}(\zeta, \zeta) + (\omega_{11,t}(\zeta + \iota, \zeta + \iota) + \omega_{12,t}(\zeta + \iota, \zeta + \iota))\varepsilon$. Since by assumption $\omega_{11,t} + \omega_{12,t} > 0$, the function $\omega_{1,t}(\zeta, \zeta)$ is monotonically increasing in ζ and therefore invertible.

changes in the spread on reserves induce substitution effects on the asset side of bank balance sheets.

Combining equations (8) and (9) for an active bank implies

$$\chi^n_{t+1} - \left(\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t\right) = \frac{1}{\eta_{n,t+1}} \frac{R^n_{t+1}}{R^f_{t+1}}, \tag{8a}$$

where we define

$$\tilde{\omega}_t(-\chi) \equiv \omega_t\left(\omega^{-1}_{1,t}(-\chi), \omega^{-1}_{1,t}(-\chi)\right) + \chi\omega^{-1}_{1,t}(-\chi).$$

Function $\tilde{\omega}_t$ summarizes how the spread on reserves, which determines reserve holdings according to condition (9), affects marginal liquidity substitution costs. Condition (8a) summarizes how the bank chooses the spread on deposits (or balance sheet length) conditional on the spread on reserves, deposit subsidies, and the deposit funding schedule.

Note from equation (8a) that the deposit rate co-moves with the risk-free interest rate but bank market power (small $\eta_{n,t+1}$) renders it "sticky." The deposit rate also reflects interest rates on reserves as well as bank leverage.[22] This is consistent with theoretical and empirical findings according to which market power and leverage (possibly constrained by capital regulation) shape the monetary policy transmission (e.g. Drechsler et al., 2017; Ulate, 2021; Wang et al., 2020). It also explains why changes in the spread on reserves do not only affect the composition of bank assets (see above) but also balance sheet length such that higher interest on reserves may drive bank loans up or down.

To see this suppose that the spread on deposits is proportional to a bank's net costs,

$$\chi^n_{t+1} \propto \nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t,$$

which holds true under plausible assumptions as we show below. Interest on reserves thus affects balance sheet length depending on three factors, namely the elasticity of $\tilde{\omega}_t$ with respect to the spread on reserves (the effect on bank costs); the markup (the effect on price, i.e., the deposit liquidity premium); and the elasticity of funding supply (the effect on quantity). At the same time, interest on reserves also affects the asset composition as we saw above, depending on the elasticity of ζ_{t+1}. When the first three factors dominate bank lending may increase in the interest rate on reserves.[23]

In the context of the equivalence analysis in section 3 we allow the central bank to extend a loan, l_{t+1}, at the gross interest rate R^l_{t+1} to banks. When the central bank posts a loan supply schedule, as we assume there, the bank's first-order condition with respect to l_{t+1} parallels the optimality condition (8a).[24]

2.3 Firms

Firms rent capital, κ_t, and labor, ℓ_t, to produce the output good. They take wages, the rental rate of capital, $R^k_t - 1 + \delta$, and the goods price as given; the rental rate reflects

[22] To see this more clearly rewrite equation (8a) as $R^f_{t+1}(1 - \nu - \tilde{\omega}_t(-\chi^r_{t+1}) + \theta_t) = R^n_{t+1}(1/\eta_{n,t+1} + 1)$ and recall that the inverse leverage ratio equals $\nu + \omega_t(\cdot) - \theta_t$.

[23] Further effects come into play when capital restrictions bind, see Bianchi and Bigio (2020).

[24] Variables l_{t+1} and R^l_{t+1} are measurable with respect to information available at date t.

the depreciation rate, δ. Without loss of generality we abstract from liquidity demand by firms. Letting f_t denote a neoclassical production function the representative firm solves[25]

$$\begin{aligned} \max_{\kappa_t, \ell_t} \quad & \pi_t^f \\ \text{s.t.} \quad & \pi_t^f = f_t(\kappa_t, \ell_t) - \kappa_t(R_t^k - 1 + \delta) - w_t \ell_t \end{aligned} \tag{10}$$

and the first-order conditions read

$$R_t^k - 1 + \delta = f_{\kappa,t}(\kappa_t, \ell_t), \tag{11}$$

$$w_t = f_{\ell,t}(\kappa_t, \ell_t). \tag{12}$$

Since f_t exhibits constant returns to scale and firms are competitive, equilibrium profits π_t^f equal zero.

2.4 Government

The consolidated government collects taxes, pays deposit subsidies, invests in capital, k_{t+1}^g, and issues money and reserves. The unit resource costs of managing money-based payments equal μ, and the unit resource costs of managing reserve-based payments among banks equal ρ. Accordingly, the government budget constraint reads[26]

$$k_{t+1}^g - m_{t+1} - r_{t+1} = k_t^g R_t^k - m_t R_t^m - r_t R_t^r + \tau_t - n_{t+1}\theta_t - m_{t+1}\mu - r_{t+1}\rho. \tag{13}$$

The government's seignorage revenue equals $m_{t+1}\chi_{t+1}^m + r_{t+1}\chi_{t+1}^r$. In a two-tier system ($m_{t+1} = 0$) and net of payment operations costs, this equals $r_{t+1}(\chi_{t+1}^r - \rho)$ which compares with $n_{t+1}\{(1 - \zeta_{t+1})\chi_{t+1}^n + \zeta_{t+1}(\chi_{t+1}^n - \chi_{t+1}^r) - \nu\}$ in the private sector. The government's share in total seignorage revenues net of payment operations costs amounts to

$$\frac{\zeta_{t+1}(\chi_{t+1}^r - \rho)}{\chi_{t+1}^n - \nu - \zeta_{t+1}\rho}.$$

Central bank liabilities are injected through open market operations. That is, banks exchange some of their capital holdings (which they acquire from households in exchange for deposits) against reserves, and households similarly exchange capital against money. In the central bank's balance sheet reserves and money thus are "backed" by capital. Alternatively, the central bank injects means of payment by transfer, without backing, reducing the central bank's capital. When a household initiates an exchange of deposits into money and the central bank accepts the incoming payment from the household's bank then the bank's reserves account at the central bank is debited or the central bank extends a loan to the bank.

[25]Variables π_t^f, κ_{t+1} and ℓ_t are measurable with respect to information available at date t.

[26]Variable k_{t+1}^g is measurable with respect to information available at date t. When the central bank extends loans to banks two additional terms are included in the constraint: l_{t+1} on the left-hand side and $l_t R_t^l$ on the right-hand side.

2.5 Market Clearing

Each household is endowed with one unit of time per period. Labor and capital market clearing as well as the bank's balance sheet identity and the definition of total profits then imply[27]

$$\ell_t = 1 - x_t, \quad \kappa_t = k_t + k_t^g + n_t - r_t, \quad \pi_t = \pi_{1,t}^b + \pi_{2,t}^b + \pi_t^f. \tag{14}$$

2.6 Resource Constraint

Walras' law implies that market clearing on the markets for labor and capital as well as the budget constraints of households, banks, firms, and the government imply market clearing on the market for the output good: Combining equations (1), (6), (7), (10), (13), and (14) yields the resource constraint

$$\kappa_{t+1} = f_t(\kappa_t, 1 - x_t) + \kappa_t(1 - \delta) - c_t - m_{t+1}\mu - n_{t+1}(\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1})) - r_{t+1}\rho. \tag{15}$$

The three right-most terms in the resource constraint are nonstandard. They represent the resource costs of payment operations of banks and the central bank as well as the liquidity substitution costs of banks. In a narrow bank regime the total resource costs of payment operations equal

$$m_{t+1}\mu + n_{t+1}(\nu + \rho)$$

because the reserves-to-deposits ratio equals unity in this case. Liquidity transformation by banks generates additional (positive or negative) resource costs

$$n_{t+1}\left(\omega_t(\zeta_{t+1}, \zeta_{t+1}) - (1 - \zeta_{t+1})\rho\right).$$

As this expression makes clear narrow banking in a two-tier payment system is resource inefficient if $\rho > -(\omega_{1,t}(1,1) + \omega_{2,t}(1,1))$.

2.7 Policy and Equilibrium

Let $\xi_{t+1} \equiv n_{t+1}/z_{t+1}$ denote the share of deposits in effective real balances. A *policy* $\mathcal{P}$ consists of

- $\{\tau_t, \theta_t\}_{t\geq 0}$;
- $\{\chi_{t+1}^r\}_{t\geq 0}$ if the central bank issues reserves; and
- $\{m_{t+1}\}_{t\geq 0}$, $\{\chi_{t+1}^m\}_{t\geq 0}$, or $\{\xi_{t+1}\}_{t\geq 0}$ if the central bank issues money and targets its quantity, spread, or share in effective real balances, respectively.

An *equilibrium* conditional on $\mathcal{P}$ consists of

- a positive allocation, $\{c_t, x_t, k_{t+1}, k_{t+1}^g, \kappa_{t+1}, \ell_t\}_{t\geq 0}$;
- positive money, deposit, and reserve holdings, $\{m_{t+1}, n_{t+1}, r_{t+1}\}_{t\geq 0}$; and

[27] When the central bank extends loans to banks the second equality reads $\kappa_t = k_t + k_t^g + n_t + l_t - r_t$.

- a positive (shadow) price system, $\{w_t, R^k_{t+1}, R^f_{t+1}, \chi^m_{t+1}, \chi^n_{t+1}, \chi^r_{t+1}\}_{t\geq 0}$,

such that (1)–(14) (and by implication (15)) are satisfied and asset markets clear. If the central bank extends loans then the policy also includes a loan funding schedule, the equilibrium objects also include $\{l_{t+1}, R^l_{t+1}\}_{t\geq 0}$, and the loan market must clear as well.

Note that multiple policies may implement the same equilibrium. For example, when money circulates the central bank may target m_{t+1} and let χ^m_{t+1} adjust to clear the money market, or it may target χ^m_{t+1} and let m_{t+1} adjust. More interestingly, when deposits circulate the government might be able to affect the premium χ^n_{t+1} equally by setting a deposit subsidy or by targeting a liquidity premium on money. In section 4 we study in detail how the government can optimally exploit instrument redundancy.

2.8 Functional Form Assumptions

Existence and uniqueness of equilibrium in models with money in the utility function may require conditions on primitives (see, e.g., Walsh, 2017). Our analysis applies for general functional forms, u, that satisfy these conditions. When closed-form solutions or simulations require more structure we impose the following functional form assumptions:

Assumption 1. Preferences satisfy

$$u(c_t, z_{t+1}, x_t) = \left((1-\vartheta)c_t^{1-\psi} + \vartheta z_{t+1}^{1-\psi}\right)^{\frac{1-\sigma}{1-\psi}} (1-\sigma)^{-1} + v(x_t),$$

where $\vartheta, \psi \in (0,1)$; $\sigma > 0, \neq 1$; and v is strictly increasing and concave.[28]

Under assumption 1 the elasticity of substitution between c_t and z_{t+1} equals ψ^{-1}.

Assumption 2. Bank resource costs due to liquidity substitution satisfy

$$\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1}) = \phi_t(1-\zeta_{t+1})^{\varphi}(1-\bar{\zeta}_{t+1})^{\bar{\varphi}},$$

where $\phi_t > 0, 1 < \varphi < \bar{\varphi}$.

In appendix A we derive this specification from micro foundations. To analyze a model without reserves layer we may set $\phi_t = 0$.

2.9 Cash and Government Bonds

Abstracting from cash and government bonds does not undermine the generality and relevance of our analysis. To see this consider cash first. Conceptually, it would be straightforward to include cash as a third central bank liability (in addition to reserves and money) and third retail means of payment (in addition to money and deposits). Cash would enter like money except that the interest rate on cash would be constrained to equal the deflation rate. Differences in liquidity between cash and money would be reflected in different λ_t values.

[28]In Niepelt (2020a) we also consider the multiplicatively rather than additively separable case.

The new insights gained from such an extended model would mainly concern cash-CBDC substitution, a swap of central bank liabilities without major macroeconomic consequences.[29] Our framework instead focuses on the "disruptive" deposit-CBDC substitution that is relevant for the banking sector and the macro economy. Cash could also be macro economically relevant by giving rise to a binding effective lower bound on interest rates, but only if prices were sticky which they are not in our setting (see section 5).[30]

Turning to government bonds, including them as a second central bank asset (in addition to capital), third bank asset (in addition to reserves and capital), and fifth household asset (in addition to capital, money, deposits, and bank equity) would be conceptually straightforward as well. But again, the insights to be gained would be limited. Government bonds as central bank asset would be irrelevant because only the consolidated government budget constraint matters. Moreover, since taxes in the model are nondistorting and households homogeneous government debt would leave the shadow value of public funds unchanged. Government bonds would also be irrelevant in household balance sheets, except possibly as instruments to hedge if risk were important and the returns on money, deposits and bonds varied differentially across contingencies, which is not the case empirically.[31]

But if bonds did not offer hedging benefits for households then banks would have no incentive to hold bonds either.[32] Bonds might be useful as bank assets, however, if they provided liquidity services, for instance as collateral in repo transactions. But in the model reserves already play that role and empirically, banks hold few treasury securities (Hanson et al., 2015, p. 452). Since both reserves and government bonds are liabilities of the consolidated government, the omission of bonds is unimportant unless we are specifically interested in the composition of banks' liquid assets.[33]

3 General Equilibrium

3.1 Model Solution

The model's equilibrium conditions reduce to three core equations. Appendix B contains detailed derivations under the CES preference assumption 1, which implies that house-

[29]For an early discussion of substitution between cash and electronic money, see BIS (1996). Keister and Sanches (2022) find minor effects of a "cash-like" CBDC.

[30]The presence of cash may also affect the market power of banks (Drechsler et al., 2017; Lagos and Zhang, 2020) and thus the elasticity of deposit funding. This mechanism operates independently of CBDC. As mentioned before, the model does not restrict the elasticity.

[31]See, for example, Divino and Orrillo (2017). See Brunnermeier et al. (2022) and Di Tella (2020), respectively, for the role of total government or central bank liabilities in the face of aggregate and idiosyncratic risks.

[32]Recall that at the margin capital and bank equity have identical risk-return characteristics. Households could replicate the effect of a bank's bond holdings on the risk-return characteristics of marginal bank equity by holding bonds themselves.

[33]If banks valued the liquidity services of government bonds and reserves in parallel to how households value liquidity services of deposits and money then the composition between reserves and bonds would be trivial: It would either be indeterminate (with both assets paying the same interest rate, controlling for liquidity differences) or in a corner.

holds hold real balances in proportion to consumption with the factor of proportionality reflecting the spread on real balances, χ_{t+1}.[34] The three core equations are

$$
\begin{aligned}
c_t^{-\sigma} &= \beta \mathbb{E}_t \left[(1-\delta+f_{\kappa,t+1}(\kappa_{t+1}, 1-x_{t+1}))c_{t+1}^{-\sigma} \frac{\Omega_{t+1}^c}{\Omega_t^c} \right], \\
v'(x_t) &= c_t^{-\sigma} f_{\ell,t}(\kappa_t, 1-x_t)\Omega_t^c, \\
\kappa_{t+1} &= f_t(\kappa_t, 1-x_t) + \kappa_t(1-\delta) - c_t \Omega_t^{rc},
\end{aligned}
$$

where Ω_t^c and Ω_t^{rc} are defined in the appendix. Ω_t^c is present because real balances affect the marginal utility of consumption, and Ω_t^{rc} accounts for the resource costs of money, deposits, and reserves. These latter costs depend on real balances relative to consumption (and thus, on the spread χ_{t+1}); payment operation and liquidity substitution costs; and the composition of real balances between money and deposits, ξ_{t+1}. In the limit where households do not value liquidity services, $\vartheta \to 0$, both Ω_t^c and Ω_t^{rc} collapse to unity and the system reduces to the baseline real business cycle model.

In steady state the inter temporal pseudo wedge, $\Omega_{t+1}^c/\Omega_t^c$, disappears and the capital-labor ratio satisfies the modified golden rule. In contrast, the pseudo wedges in the resource constraint and the intra temporal first-order condition remain relevant even in steady state because real balances generate resource costs and have asymmetric effects on the marginal utilities of consumption and leisure.[35]

3.2 Monetary Policy Transmission

In a money based system the central bank directly determines the spread χ_{t+1} and it directly controls all pseudo wedges; in appendix B we characterize the implied equilibrium under the CES preference assumption 1. In a two-tier system, in contrast, the transmission mechanism of monetary policy operates through the cost structure of banks and their price setting. The spread on reserves influences the reserves-to-deposits ratio of banks; this pins down their portfolio returns and the costs of liquidity substitution, both directly and indirectly (due to externalities); and these costs together with deposit subsidies determine the liquidity premium that banks charge depositors.

Under preference assumption 1 banks charge a constant markup over costs,

$$
\chi_{t+1}^n = \frac{\nu + \tilde{\omega}_t(-\chi_{t+1}^r) - \theta_t}{1-\psi},
$$

a result we had anticipated when discussing the effect of χ_{t+1}^r on bank balance sheets. As the condition makes clear, both a higher deposit subsidy and a lower liquidity premium on reserves raise the equilibrium deposit rate. A lower liquidity premium raises the deposit rate at less fiscal cost than a higher deposit subsidy: While a marginal increase

[34]The spread on real balances equals χ_{t+1}^n when households hold deposits; χ_{t+1}^m/λ_t when they hold money; and $\chi_{t+1}^n = \chi_{t+1}^m/\lambda_t$ when both means of payment circulate.

[35]The economy exhibits monetary neutrality. When all spreads are orthogonal to inflation it also exhibits superneutrality. See Walsh (2017, 2.4.2) for a textbook treatment of nonsuperneutrality in the Sidrauski (1967) model.

in θ_t imposes fiscal costs n_{t+1} and lowers χ^n_{t+1} by $(1-\psi)^{-1}$, a marginal increase in the interest rate on reserves (corresponding to a marginal increase in $-\chi^r_{t+1}$) costs $n_{t+1}\zeta_{t+1}$ and increases χ^n_{t+1} by $\tilde{\omega}'_t \cdot (1-\psi)^{-1}$, which is negative and smaller than $-\zeta_{t+1}(1-\psi)^{-1}$.[36] This is a consequence of the envelope theorem and the fact that reserves generate externalities. In appendix B we further characterize the equilibrium in a two-tier system under assumption 2 about the liquidity substitution costs, $\tilde{\omega}_t(-\chi^r_{t+1})$.

We also characterize the equilibrium in a mixed system with deposits and money. In this case, the policy mix governing the liquidity supply is critical. If policy targets m_{t+1}, it affects the elasticity of deposit funding and modifies banks' price setting. If policy targets the composition of real balances, ξ_{t+1}, it affects bank profits but not their price setting. And if policy targets χ^m_{t+1} it can force banks to set the deposit spread accordingly or drive them out of business.[37] We defer a more detailed discussion of these policy mixes to the normative analysis in section 4.

Calibrated Example We compute impulse response functions under assumptions 1 and 2 using the sequence-space Jacobian approach developed in Auclert et al. (2021).[38] Appendix C describes the calibration, which mainly uses information on bank net interest margins and average annual asset returns (FDIC); estimates of the expenditure share for liquidity services (Di Tella, 2020); estimates of the costs for banks and central banks of managing retail payments (Schmiedel et al., 2012); and fire sale prices (Shleifer and Vishny, 2011). Table 1 in the appendix summarizes the calibrated parameter values.

Based on this calibration we simulate the effects of an increase in the liquidity premium on reserves, χ^r_{t+1}, corresponding to a reduction in R^r_{t+1} relative to R^f_{t+1}. The change amounts to 25 basis points on an annual basis and lasts for twelve quarters. Figure 3 in appendix D illustrates how the intervention affects interest rates and bank choices: On impact, the interest rate on reserves falls as does the interest rate on deposits. Banks' reserve holdings and households' deposit holdings fall and banks' costs of liquidity substitution rise. Figure 1 illustrates the effects on the allocation: The lower deposit rate induces a substitution effect in favor of consumption relative to deposits. Moreover, the change in deposit holdings affects the marginal utility of consumption; when $\sigma > \psi$, as the calibration stipulates, lower deposit holdings increase the marginal utility of consumption. As a consequence, consumption rises while income effects increase labor supply.

The interest rate on capital and thus the risk-free rate increase slightly. Accordingly, consumption and leisure grow during the first 12 quarters. The capital stock shrinks as investment falls short of its steady-state value and output exceeds its steady-state value, due to the stronger labor supply and in spite of the depleted capital stock. When the liquidity premium is normalized after 12 quarters investment rebounds and all variables

[36] Note that $\tilde{\omega}'_t = (\omega_{1,t} + \omega_{2,t}) \cdot \zeta'_{t+1} - \zeta_{t+1} + \chi^r_{t+1}\zeta'_{t+1} = \omega_{2,t} \cdot \zeta'_{t+1} - \zeta_{t+1} < -\zeta_{t+1}$ where we use the bank's first-order condition.

[37] A very high liquidity premium on money, which renders money unattractive relative to deposits, has no effects.

[38] A Jupyter notebook with Python code is available on request. The code is based on software developed by Adrien Auclert, Bence Bardóczy, Michael Cai and Matthew Rognlie, downloadable at https://github.com/shade-econ/sequence-jacobian.

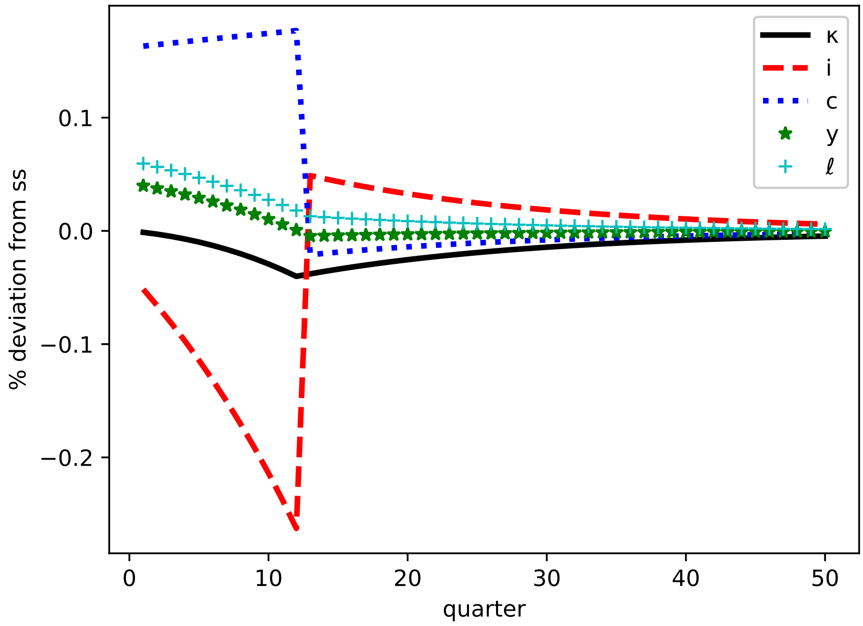


Figure 1: Responses to an increase of χ^r_{t+1} by 25 basis points (annual) at $t = 1$ that lasts for twelve quarters.

gradually revert to their steady-state values. The example illustrates how a reduction in the policy rate (policy easing) may depress investment while simultaneously stimulating consumption and output.

In appendix D we also present the impulse response functions for a shock to the subsidy rate, θ_t. A persistent increase in θ_t raises R^n_{t+1}, r_{t+1}, and n_{t+1} while it has no effects on ω_t. Qualitatively, the effects on the allocation are similar to those in the case of an increase in χ^r_{t+1} (discussed above), but with the opposite sign. The example illustrates how deposit subsidies can stimulate investment while depressing labor supply, consumption, and output.

Changes in the calibration alter the magnitude of these effects. A smaller (higher) value for ψ (ϑ) renders the effects on the allocation more pronounced, and in the case of ψ it also noticeably strengthens the effects on reserves and deposits. A higher labor supply elasticity strengthens the responses on the labor market, and this feeds into consumption and output. A higher elasticity of the costs of liquidity substitution (φ) increases the consumption and investment responses. In the case of the χ^r_{t+1} shock the interest rate on deposits as well as reserves and deposits fall by less than in the baseline; in the case of the θ_t shock the effect on reserves and deposits strengthens.

The sign of $\sigma - \psi$ affects the responses qualitatively. With $\sigma < \psi$ (in contrast to what the calibration stipulates) lower deposit holdings decrease the marginal utility of consumption and this Pareto complementarity reverses the sign of the effects on the allocation. For instance, in response to the hike in χ^r_{t+1}, investment rises while consumption, labor supply, and output fall. This reminds of the important role of Pareto complementarity vs. substitutability in monetary models.[39]

3.3 Equivalence

In an environment with no reserves and no resource costs of liquidity provision Brunnermeier and Niepelt (2019) show that public money (e.g., CBDC) can replace private money (e.g., deposits) without altering the equilibrium allocation or the price system. We extend this equivalence result to our environment with reserves and resource costs: As long as the public and the private sector provide liquidity equally efficiently the central bank can always ensure that a portfolio shift from deposits into money leaves equilibrium consumption, capital accumulation, and the price system unchanged. Private and public means of payment thus are substitutes in general equilibrium and portfolio shifts out of deposits into CBDC do not undermine bank intermediation as long as the central bank intervenes appropriately.

The equivalence result follows under a parameter restriction but otherwise under much more general conditions than those laid out in section 2 as we explain below.[40] The parameter restriction stipulates that the resource costs per unit of effective real balances are the same for money and deposits:

Condition 1. $\mu/\lambda_t = \nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) + \zeta_{t+1}\rho$ where ζ_{t+1} denotes the equilibrium reserves-to-deposits ratio.

The necessity of condition 1 is clear: Different unit resource costs of money and deposits would necessarily undermine equivalence because they would alter resource requirements and thus the allocation. A special case where the condition is satisfied arises when public and private payment operations in a narrow banking system generate commensurate costs, $\mu/\lambda_t = \nu + \rho$, and liquidity transformation is costless, $\omega_t(\zeta_{t+1}, \zeta_{t+1}) = (1 - \zeta_{t+1})\rho$.[41] More generally, the condition requires cost disadvantages of narrow banks relative to public money provision to be compensated by cost savings of liquidity transformation.

The following result is formally stated and proved in appendix E:

Proposition 1. Suppose condition 1 holds. Consider a policy that implements an equilibrium with deposits and reserves. There exists another policy and equilibrium with fewer deposits and reserves, more money, a central bank loan, a different ownership structure of capital, and otherwise the same allocation and price system.[42] The central bank loan carries the interest rate

$$R^l_{t+1} = \frac{R^n_{t+1} + (\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t)R^f_{t+1} - \zeta_{t+1}R^r_{t+1}}{1 - \zeta_{t+1}}.$$

[39]See, for instance, Wang and Yip (1992).

[40]For added generality we allow for stochastic interest rates on all means of payment: Variables $R^r_{t+1}, R^m_{t+1}, R^n_{t+1}$ are measurable with respect to information available at date $t + 1$.

[41]This is trivially satisfied when liquidity substitution is costless and $\mu = \nu = \rho = 0$.

[42]The new policy may also include state contingent taxes whose market value equals zero.

The logic underlying the proposition is simple: When households transfer one dollar from the bank to the central bank, converting deposits into money, the central bank can pass its new funding back to banks as loans. Unlike in the case of a cash run where the central bank must react timely and physically deliver cash to avert a bank collapse, pass through funding does not impose any information requirements; it occurs automatically. The bank pays the share ζ_{t+1} of the transfer out of reserves and the remaining share is financed mechanically by the central bank when it accepts the incoming payment.[43] The transfer does not put bank funding at risk or undermine financial stability.

To also preserve bank funding costs the central bank must price the loan appropriately. By posting a commensurate loan supply schedule the central bank can perfectly replicate the choice sets that banks had prior to the intervention. The central bank thus can guarantee that bank lending remains unchanged and more generally, that the initial equilibrium prices and allocation together with the modified portfolios constitute an equilibrium. Establishing these results requires an analysis of the budget and choice sets of banks, households, firms, and the government, which we relegate to appendix E.

To interpret the expression for the equivalent loan interest rate, R^l_{t+1}, note that a bank only invests the share $1-\zeta_{t+1}$ of deposit funding in capital and the rest in reserves while a central bank loan is fully invested in capital. The loan interest rate therefore reflects the (old and new) equilibrium reserves-to-deposits ratio, ζ_{t+1}. A higher ratio increases R^l_{t+1} because less loan than deposit funding is needed (reflected by the term $1-\zeta_{t+1}$ in the denominator); it decreases R^l_{t+1} because reserves reduce the costs of liquidity substitution (the $\omega_t(\zeta_{t+1}, \zeta_{t+1})$ term); and it affects R^l_{t+1} through a third channel when the interest rates on deposits and reserves differ (the term $(R^n_{t+1} - \zeta_{t+1}R^r_{t+1})/(1-\zeta_{t+1})$).

Appendix E offers a detailed discussion of the interpretation of proposition 1 as well as its robustness and possible sources of nonequivalence. Here we provide a short summary. First, the equivalence result states that an equilibrium allocation and price system is associated with multiple compositions of real balances cum policies. This multiplicity is distinct from the instrument redundancy discussed elsewhere according to which multiple policies may implement the same equilibrium including a specific portfolio structure. Second, the result implies that under condition 1 welfare in an economy with deposits is weakly lower than in an optimally managed economy with CBDC (and deposits). Third, an introduction of CBDC would require smaller central bank loans today than before the financial crisis.

Fourth, the result holds for general preferences, monetary frictions, market structures, and initial equilibria and it also applies in the presence of bank runs. Neither fixed costs of payment networks nor household or bank heterogeneity need to undermine the result.

Finally, the proof of proposition 1 points to possible causes of nonequivalence some of which the literature has touched upon: Condition 1 may be violated; nonlinear substitutability of money and deposits in household real balances may undermine "liquidity neutrality" (Brunnermeier and Niepelt, 2019); information limitations may prevent the central bank from posting the appropriate loan schedule (Niepelt, 2020b); the central

[43] Accepting the incoming payment goes hand in hand with acquiring a claim against the bank, by double entry book keeping. When the central bank rejects the incoming payment then the bank balance sheet does not change.

bank may inject CBDC by transfer, thereby affecting the market for real balances, as discussed by Keister and Sanches (2022, 6); or it may issue CBDC only in exchange for bonds, as in Kumhof and Noone (2021) and Barrdear and Kumhof (2022).[44]

Moreover, deposits and central bank loans may affect the choice sets of banks differently (aside from their role as sources of funding), for instance because of different collateral requirements (Williamson, 2019; Böser and Gersbach, 2020).[45] In the appendix we explain why this does not imply that the equivalence result prescribes an "unrealistic" policy; it rather points to a potential inconsistency of central bank policies. Or, gross balance sheet positions may directly affect resource costs, for example because of costly central bank asset management. Piazzesi and Schneider (2021) consider an environment that includes such elements.

3.4 Bank Funding-Cost-Reduction-at-Risk

Under the equivalent policy characterized in proposition 1 the central bank would lengthen its balance sheet (unless $\lambda_t^{-1} < \zeta_{t+1}$, see appendix E) and replace deposit taking banks as liquidity providers for the private sector. This could change how conflicts between interest groups and within the government are resolved.[46] As a consequence the equilibrium policy after the introduction of CBDC could differ from the equivalent one and the allocation could change even if condition 1 held.

Suppose that after the introduction of CBDC the central bank refinanced banks at the risk-free interest rate, R^f_{t+1}, rather than the equivalent loan rate, R^l_{t+1}. This would increase bank funding costs as long as $R^f_{t+1} > R^l_{t+1}$, and if banks had to fund all capital investment out of loans rather than deposits their costs would rise in proportion to $n_{t+1}(1 - \zeta_{t+1})$. We refer to the cost increase relative to GDP as the "funding-cost-reduction-at-risk,"

$$\text{fcr}_t \equiv \frac{R^f_{t+1} - R^l_{t+1}}{R^f_{t+1}} \frac{n_{t+1}(1 - \zeta_{t+1})}{\text{GDP}_t}.$$

In appendix F we compute fcr_t for U.S. banks following different strategies. Our preferred approach uses an equilibrium condition of the model under assumption 1 and

[44]Other factors are negative effects on innovation from public rather than private liquidity provision; changes in the asset span (Brunnermeier and Niepelt, 2019) or market participation (Benigno et al., 2022); restrictions on admissible taxes; or political economy factors, see below.

[45]Synergies between bank assets and liabilities such as in Kashyap et al. (2002) or Hanson et al. (2015) do not undermine equivalence as long as the synergies are also present with a central bank loan. Pulley and Humphrey (1993) offer an empirical assessment of such synergies.

[46]For example, a transparent single-tier system with retail liquidity provision by the central bank could strengthen political opposition against "too cheap" bank funding. Or, the treasury could subject the central bank to more intense fiscal pressure once the change to a single-tier system linked public sector seignorage more closely to real balances. A rigorous analysis would require a framework with heterogeneous households, political aggregation of preferences, as well as strategic fiscal and monetary authorities with conflicting interests. For a brief discussion see Niepelt (2021).

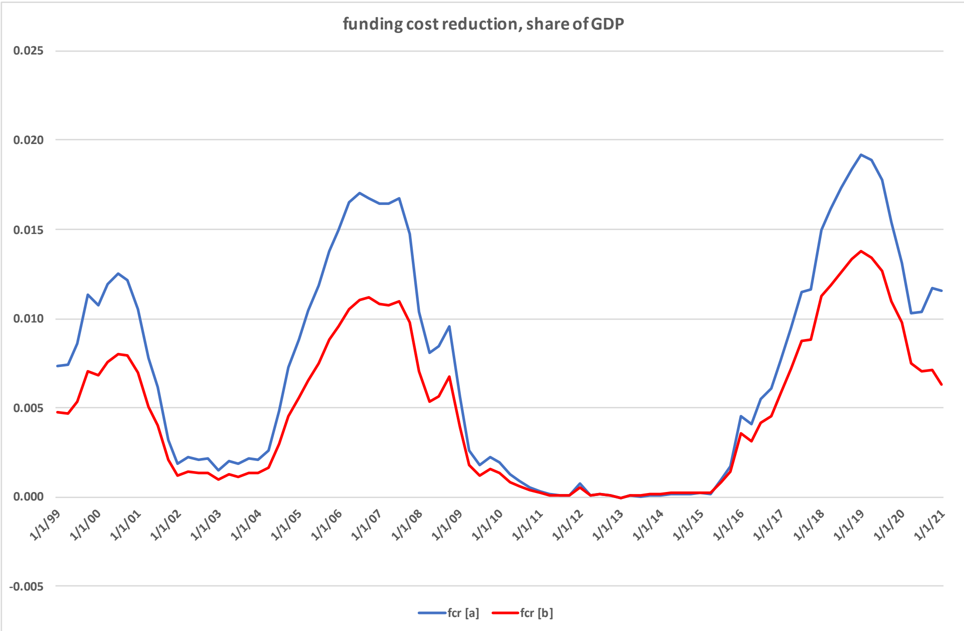


Figure 2: Two measures of banks' funding-cost-reduction-at-risk under assumption 1.

yields the simple expressions

$$\begin{aligned} R^l_{t+1} &= R^f_{t+1} - \frac{\psi}{1-\zeta_{t+1}}(R^f_{t+1} - R^n_{t+1}), \\ \mathrm{fcr}_t &= \psi \chi^n_{t+1} \frac{n_{t+1}}{\mathrm{GDP}_t}, \end{aligned}$$

according to which bank market power (ψ) reduces the equivalent loan rate and thus increases the funding-cost-reduction-at-risk. In addition to market power only two statistics are required to calculate fcr_t: The deposit spread and the deposits-to-GDP ratio.

Figure 2 illustrates the time series for fcr_t since 1999 for two alternative measures of n_{t+1}: The funding-cost-reduction-at-risk varies substantially, is always positive (this changes when we follow an alternative strategy, see the appendix), and at times exceeds 1.5 percent of GDP.[47] In appendix F we offer a detailed discussion.

4 Optimality

We have seen that CBDC has real effects when condition 1 is violated. This raises the questions, which type of monetary system is preferable, and which policy conditional on a given system. To address these questions we first characterize the allocation chosen by a social planner that is constrained by the production and payment technologies introduced

[47] These numbers compare with NIPA data for financial sector profits on the order of 3 percent of GDP prior to the financial crisis, negative profits during the crisis, and 2 to 3 percent after the crisis.

in section 2.[48] Thereafter, we analyze whether the Ramsey government can implement the planner allocation. We also consider the implications of admissibility restrictions on policy instruments, tax distortions, and too-big-to-fail banks.

4.1 Social Planner Allocation

The social planner solves

$$\begin{aligned}
&\max_{\{c_t, x_t, \kappa_{t+1}, m_{t+1}, n_{t+1}, r_{t+1}\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}_0 \left[u(c_t, z_{t+1}, x_t) \right] \\
\text{s.t.} \quad & \kappa_{t+1} = f_t(\kappa_t, 1 - x_t) + \kappa_t(1-\delta) - c_t - m_{t+1}\mu - n_{t+1}(\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1})) - r_{t+1}\rho, \\
& \kappa_{t+1}, m_{t+1}, n_{t+1}, r_{t+1} \geq 0.
\end{aligned}$$

The optimality conditions for capital, consumption, and leisure are standard and parallel the conditions in the decentralized equilibrium (see equations (2), (3), (11), (12), and (14)). To minimize the costs of liquidity transformation the planner issues reserves until their marginal costs equal the benefits from reducing banks' liquidity substitution costs,

$$\begin{aligned}
\omega_{1,t}(\zeta_{t+1}, \zeta_{t+1}) + \omega_{2,t}(\zeta_{t+1}, \zeta_{t+1}) + \rho &= 0 \ \text{ if } \ n_{t+1} > 0, \\
r_{t+1} &= 0 \ \text{ otherwise.}
\end{aligned}$$

When reserves generate externalities ($\omega_{2,t} \neq 0$) the planner takes this into account—unlike a bank in decentralized equilibrium (see equation (9)). The curvature of the ω_t function implies a unique mapping from ρ to the optimal ζ_{t+1} choice.[49] Finally, the first-order conditions for money and deposits yield the optimality condition

$$u_z(c_t, z_{t+1}, x_t) = u_c(c_t, z_{t+1}, x_t) \min[\mu/\lambda_t, \nu + \omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1}) + \zeta^{\star}_{t+1}\rho], \tag{SP}$$

where we use the optimality condition for reserves and $\zeta^{\star}_{t+1}$ denotes the optimal reserves-to-deposits ratio in a two-tier system.

To interpret equation (SP), consider first the case without a reserves layer, $\omega_t = 0$. In this case, condition (SP) reduces to $u_z(c_t, z_{t+1}, x_t) = u_c(c_t, z_{t+1}, x_t) \min[\mu/\lambda_t, \nu]$ and the planner provides real balances up to the point where the marginal liquidity benefit equals the marginal social costs. This is just a variant of the Friedman (1969) rule which, in its common form, abstracts from resource costs of managing liquidity, i.e., posits $\min[\mu/\lambda_t, \nu] = 0$. As for the source of liquidity, the planner generically provides one means of payment (but see the discussion below). When λ_t strictly exceeds μ/ν then the first best involves money but no deposits, and vice versa. Only when the relative

[48] A "true" social planner, of course, would not require liquidity to implement an allocation. Our social planner construct amounts to a government without instrument admissibility constraints.

[49] By the mean value theorem, for any $\varepsilon > 0$ there exists a $\iota \in (0, \varepsilon)$ such that $\omega_{1,t}(\zeta+\varepsilon, \zeta+\varepsilon) + \omega_{2,t}(\zeta+\varepsilon, \zeta+\varepsilon) = \omega_{1,t}(\zeta, \zeta) + \omega_{2,t}(\zeta, \zeta) + (\omega_{11,t}(\zeta+\iota, \zeta+\iota) + 2\omega_{12,t}(\zeta+\iota, \zeta+\iota) + \omega_{22,t}(\zeta+\iota, \zeta+\iota))\varepsilon$. Since by assumption $\omega_{11,t} + \omega_{12,t} > 0$ and $\omega_{21,t} + \omega_{22,t} > 0$, the function $\omega_{1,t}(\zeta, \zeta) + \omega_{2,t}(\zeta, \zeta)$ is monotonically increasing in ζ and therefore invertible.

costs of managing money- and deposit-based payments and the relative liquidity benefits happen to coincide is the planner indifferent between the two payment systems.

Consider next the case of interest in which liquidity transformation requires banks to engage in liquidity substitution. The new element in equation (SP) then is that the planner also accounts for the indirect costs of a two-tier system, namely those related to reserve-based payments and banks' liquidity substitution. Summarizing:

Proposition 2. The social planner provides the total-cost-minimizing means of payment. It equalizes the marginal liquidity benefit of real balances and the marginal social costs.

Versions of proposition 2 hold under general conditions. In particular, our conclusions would not be affected if money or deposits entered nonlinearly in effective real balances or in the costs of operating payments, such that both money and deposits optimally circulated.[50] In general, the "total-cost-minimizing means of payment" is a combination of money and deposits.

4.2 Ramsey Policy

Unlike the social planner, the Ramsey government controls the allocation only indirectly, by choosing a feasible policy $\mathcal{P}$. We show next that the Ramsey policy $\mathcal{P}^{\star}$ supports an equilibrium with the social planner allocation. Since any equilibrium satisfies the social planner's optimality conditions for capital, consumption, and leisure as well as the resource constraint, the Ramsey policy implements the first best if the first-best quantities of money, deposits, and reserves correspond to the demand and supply of these means of payment under $\mathcal{P}^{\star}$.

Suppose first that the social planner uses money rather than deposits, $\mu/\lambda_t < \nu + \omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1}) + \zeta^{\star}_{t+1}\rho$. The Ramsey policy then implements the first-best by issuing the efficient amount of money, charging the liquidity premium

$$\chi^{m\star}_{t+1} \equiv \mu, \tag{RA-1}$$

and pricing banks out of the market. Equation (RA-1), which reduces to the traditional Friedman rule when $\mu = 0$, follows directly from equations (4) and (SP). Households would only hold deposits if $\chi^{n}_{t+1} \leq \chi^{m}_{t+1}/\lambda_t = \mu/\lambda_t$ (see equation (5)) but such a low deposit premium would generate losses if the government set subsidies to zero and increased the liquidity premium on reserves sufficiently.

Suppose next that the planner relies on deposits rather than money. Consider a relaxed Ramsey program without the bank's optimality condition for deposits, equation (8). In this relaxed program the government implements the first best by issuing the efficient quantity of deposits (and no money), charging the liquidity premium

$$\chi^{n\star}_{t+1} \equiv \nu + \omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1}) + \zeta^{\star}_{t+1}\rho, \tag{RA-2}$$

[50] For example, if real balances were given by $z_{t+1} \equiv \Lambda_t(m_{t+1}) + n_{t+1}$ with $\Lambda_t(m_{t+1}) \equiv A\ln(m_{t+1}+1)$, the first-order conditions would hold as before except that λ_t would be replaced by $A/(m_{t+1}+1)$. For $A > \mu/(\nu+\omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1})+\zeta^{\star}_{t+1}\rho)$ but not "too large" the planner would rely on both money and deposits (and reserves).

and setting the liquidity premium on reserves to

$$\chi^{r\star}_{t+1} \equiv -\omega_{1,t}(\zeta^\star_{t+1}, \zeta^\star_{t+1}). \tag{RA-3}$$

Equation (RA-2) again follows from equations (4) and (SP). The liquidity premium on reserves in (RA-3) follows from equation (9) and the planner's optimal choice of reserves; it induces banks to select the first-best reserves-to-deposits ratio even in the presence of externalities. We conclude that the allocation in the relaxed program is first-best.[51]

In the full Ramsey program the additional equilibrium constraint (8) or (8a) does not reduce the choice set as long as the government can flexibly employ the deposit subsidy. The appropriate θ_t choice induces banks to charge the optimal liquidity premium on deposits given in (RA-2) conditional on the optimal liquidity premium on reserves given in (RA-3). From equation (8a) it is given by

$$\theta^\star_t \equiv \frac{1}{\eta_{n,t+1}} \frac{R^{n\star}_{t+1}}{R^{f\star}_{t+1}} - \chi^{n\star}_{t+1} + \nu + \tilde{\omega}_t(-\chi^{r\star}_{t+1}),$$

where $R^{f\star}_{t+1}$ is pinned down by the first-best allocation and $R^{n\star}_{t+1} \equiv R^{f\star}_{t+1}(1 - \chi^{n\star}_{t+1})$.[52] Using (RA-2), (RA-3), and bank optimality this can be expressed as

$$\theta^\star_t = \frac{1}{\eta_{n,t+1}} \frac{R^{n\star}_{t+1}}{R^{f\star}_{t+1}} + \zeta^\star_{t+1}\omega_{2,t}(\zeta^\star_{t+1}, \zeta^\star_{t+1}). \tag{RA-4}$$

The optimal subsidy has two components, reflecting the two frictions in the banking sector. Stronger market power (a lower elasticity of deposit funding) requires a higher subsidy as banks must be encouraged to expand their balance sheets. More surprisingly, a stronger reserves externality demands the opposite (recall that $\omega_{2,t} < 0$). Intuitively, when reserves generate external benefits the Ramsey policy increases their interest rate, and to sterilize the effect of this subsidy on the deposit margin it lowers the deposit subsidy. Summarizing:

Proposition 3. The Ramsey policy implements the first best independently of whether the social planner relies on money or deposits. In the former case the Ramsey policy satisfies (RA-1). In the latter case it satisfies (RA-3) and (RA-4) and implements the liquidity premium given in (RA-2); deposits may be taxed or subsidized.

Again, versions of the proposition hold even when λ_t is endogenous because the policy instruments still suffice to control liquidity, its composition, and banks' willingness to issue deposits.

When optimality is consistent with the joint circulation of money and deposits, i.e., when the condition $\mu/\lambda_t = \nu + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \zeta^\star_{t+1}\rho$ happens to be satisfied, the Ramsey

[51] In the model without a reserves layer the right-hand side of equation (RA-2) is replaced by ν; the Ramsey government need not target the reserves-to-deposits ratio; and it lacks one instrument, χ^r_{t+1}. The first best can thus be implemented.

[52] In general $\mathcal{P}^\star$ may not uniquely implement the first best. When we impose functional form assumptions this is not an issue, see appendix H.

government may issue both money and reserves. Proposition 3 indicates that CBDC and reserves should be remunerated at different rates in this case, reflecting the costs of payment operations on the one hand and of liquidity substitution subject to externalities on the other. To implement the optimal rates R^m_{t+1} and R^r_{t+1} thus requires that the government can price discriminate between wholesale and retail users of central bank liabilities.

Under functional form assumptions we can solve for simple Ramsey policy rules (see appendix H). Under assumption 2 the optimal liquidity premium on reserves equals

$$\chi^{r\star}_{t+1} = \rho \frac{\varphi}{\varphi + \bar{\varphi}} \leq \rho,$$

which falls short of the operating costs of reserve-based payments when reserve holdings generate externalities. For $\rho \to 0$ reserves optimally exhibit no liquidity premium and the optimal liquidity premium on deposits approaches ν.

Under assumption 1 the optimal subsidy equals

$$\theta^\star_t = \psi \left(\nu + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \zeta^\star_{t+1}\rho\right) - \zeta^\star_{t+1}(\rho - \chi^{r\star}_{t+1}).$$

The rate equals zero when there is no market power ($\psi \to 0$) and reserves are not subsidized ($\chi^{r\star}_{t+1} = \rho$), or when $\psi > 0$ and $\chi^{r\star}_{t+1} < \rho$ but the two effects exactly cancel. Deposits are taxed when banks have limited market power and reserves generate substantial externalities. We discuss additional results in appendix H.

When we impose the calibrated parameter values we find (on an annual basis)

$$\chi^{r\star}_{t+1} = 0.0004, \quad \theta^\star_t = 0.0018, \quad \chi^{n\star}_{t+1} = 0.0072.$$

That is, reserves should pay nearly the risk-free interest rate; deposits should (still) be subsidized; and deposits should pay roughly seventy basis points less than the risk-free rate. In a CBDC system the interest rate on money should be slightly higher, $\chi^{m\star}_{t+1} = 0.0077$. In other words, operating a two- rather than single-tier payment system requires slightly fewer resources (under the assumption that $\lambda_t = 1$), $\mu/\lambda_t > \nu + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \zeta^\star_{t+1}\rho$. The cost advantage of the two-tier system equals $1 - \chi^{n\star}_{t+1}/\chi^{m\star}_{t+1}$ or roughly seven percent.

4.3 Instrument Restrictions and Instrument Redundancy

In addition to the spread on reserves—a common monetary policy instrument—the Ramsey policy analyzed so far relies on a—less common—deposit subsidy, θ_t, (or similar instruments such as regulatory constraints) to shape bank balance sheets. What happens when such an instrument is not admissible?

To fix ideas suppose that the planner relies on deposits rather than money, and reserves generate externalities. The unconstrained Ramsey policy sets the liquidity premium on reserves to $-\omega_{1,t}(\zeta^\star_{t+1}, \zeta^\star_{t+1})$ and the deposit subsidy to $\theta^\star_t$ (see equations (RA-3) and (RA-4)). But if the θ_t instrument is not admissible an alternative is needed to induce banks to charge the optimal deposit spread, $\chi^{n\star}_{t+1}$. To implement the first best this alternative should not waste resources, i.e., it should not involve money issuance because that would require households to hold money rather than more cost effective deposits. But targeting the premium, χ^m_{t+1}, without actually issuing money has the potential to implement the first best.

From equation (8a) a bank whose deposits are not subsidized or taxed sets the liquidity premium as a markup over costs, at the "monopsony premium"

$$\chi^n_{t+1} = \nu + \tilde{\omega}_t(-\chi^{r\star}_{t+1}) + \frac{1}{\eta_{n,t+1}} \frac{R^n_{t+1}}{R^f_{t+1}},$$

which collapses to costs when the elasticity $\eta_{n,t+1}$ approaches infinity. When the government sets a premium on money at a level below λ_t times the monopsony premium then banks have two options: They may raise the deposit rate to the competitive level in equation (5) (plus epsilon), pricing the central bank out of the market, or they may exit. Banks choose the former option if this yields nonnegative profits.[53] Setting $\chi^m_{t+1} = \lambda_t \chi^{n\star}_{t+1}$ therefore implements the first best as long as

$$\nu + \tilde{\omega}_t(-\chi^{r\star}_{t+1}) \le \chi^{n\star}_{t+1} < \nu + \tilde{\omega}_t(-\chi^{r\star}_{t+1}) + \frac{1}{\eta_{n,t+1}} \frac{R^n_{t+1}}{R^f_{t+1}}.$$

Since the monopsony premium exceeds $\chi^{n\star}_{t+1}$ when $\theta^\star_t > 0$ (and vice versa) we have the following result:

Proposition 4. When the social planner relies on deposits, targeting χ^m_{t+1} (in addition to χ^r_{t+1}) can substitute for the first-best deposit subsidy, $\theta^\star_t$, but only if $\theta^\star_t > 0$.

Intuitively, when $\theta^\star_t > 0$ a χ^m_{t+1} target works for the same reason as in Andolfatto (2021): By offering a high interest rate on money the central bank pushes deposit rates up. When $\theta^\star_t < 0$, in contrast, monetary policy is powerless because it cannot push deposit rates down. This latter case never arises in Andolfatto (2021) because that model does not feature reserves that generate externalities.

What is the second-best policy when $\theta^\star_t < 0$ but the θ_t instrument is not admissible? CBDC offers several candidate policies including a "ξ policy" that targets the composition of real balances, $\xi_{t+1} \equiv n_{t+1}/z_{t+1}$, or an "$m$ policy" that targets the quantity of money. Either policy also targets χ^r_{t+1} and lets banks issue their preferred quantity of deposits such that spreads adjust according to condition (5). Being second best these policies generically set $\chi^r_{t+1} \neq \chi^{r\star}_{t+1}$. For example, the specific m policy with $m_{t+1} = 0$ sets the liquidity premium on reserves higher (lower) than $\chi^{r\star}_{t+1}$ when $\theta^\star_t < 0$ (> 0). m and ξ policies implement different equilibria because of their unequal effects on bank markups; which policies are optimal depends on parameters.[54]

4.4 Frictions

From instrument restrictions we turn to additional frictions in the economic environment.

[53] See also the discussion in appendix B.

[54] See appendix B and Niepelt (2020a).

Too-Big-To-Fail Banks The first friction concerns unobservable bank choices and commitment problems that give rise to bailouts of too-big-to-fail banks. Unlike in the main model, we assume that banks may invest in two types of capital and that their choice is unobserved. While both types have identical return characteristics the newly introduced type generates private benefits for bank management but has a lower price during fire sales, rendering it inefficient: The private gains are smaller than what it costs to engage in additional precautionary liquidity substitution, $\hat{\gamma}$ per unit of deposit.

We assume that the appropriate amount of liquidity substitution is not enforceable because the government observes ζ_{t+1} but cannot discern whether banks bear liquidity substitution costs $\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1})$ or $\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1}) + \hat{\gamma}$. We also assume that the government cannot commit to let banks fail when they take insufficient precautions because that would have large social costs. In equilibrium, banks therefore choose to invest in the capital with private benefits but do not pay the additional costs $\hat{\gamma}$ and this forces the government to bail banks out at social costs $\gamma \geq \hat{\gamma}$ per unit of deposit.

If γ is sufficiently large the social cost-benefit analysis favors CBDC even in situations in which the social planner opts for deposits. Formally, this situation arises when $0 < \mu/\lambda_t - (\nu + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \zeta^\star_{t+1}\rho) < \gamma$. To address the problem it does not suffice to use CBDC as a threat, as in the case with a nonadmissible θ_t instrument. Instead, the government must crowd out banks' liquidity provision and issue CBDC. Such a policy falls short of implementing the first best, however; efficiency requires that either the information or the commitment problem be resolved. Summarizing:

Proposition 5. With sufficiently large γ, information or commitment frictions in the face of too-big-to-fail banks provide a rationale for circulating CBDC even when the social planner opts for deposits.

Tax Distortions and Regulatory Costs The second friction concerns tax distortions. Unlike in the main model, we assume that taxing households (τ_t) causes deadweight burdens such that ceteris paribus, lower taxes increase welfare. To compare the required tax revenues and implied tax distortions in the single- and two-tier system we contrast government revenues from sources other than taxes, i.e., central bank profits.

In a single-tier system the central bank's profit between dates t and $t+1$ equals

$$m_{t+1}(\mathbb{E}_t[\mathrm{sdf}_{t+1}(R^k_{t+1} - R^m_{t+1})] - \mu) = m_{t+1}\left(\chi^m_{t+1} - \mu\right),$$

which collapses to zero under the Ramsey policy (see equation (RA-1)). In a two-tier system, the central bank's profit is given by

$$n_{t+1}\left\{\zeta_{t+1}(\mathbb{E}_t[\mathrm{sdf}_{t+1}(R^k_{t+1} - R^r_{t+1})] - \rho) - \theta_t\right\} = n_{t+1}\left\{\zeta_{t+1}\left(\chi^r_{t+1} - \rho\right) - \theta_t\right\}.$$

Here, two components structurally undermine budget balance under the Ramsey policy: The central bank's reserves subsidy, $\chi^{r\star}_{t+1} - \rho = \omega_{2,t}(\zeta^\star_{t+1}, \zeta^\star_{t+1}) < 0$, and the central bank's (or treasury's) deposit subsidy, $\theta^\star_t$. Summing the two components yields

$$\zeta^\star_{t+1}\left(\chi^{r\star}_{t+1} - \rho\right) - \theta^\star_t = -\frac{1}{\eta_{n,t+1}}\frac{R^{n\star}_{t+1}}{R^{f\star}_{t+1}} < 0;$$

that is, the total budgetary impact under the Ramsey policy is unambiguously negative unless banks are competitive. Intuitively, the loss from subsidizing reserves is fully balanced by one component of the optimal subsidy, $\theta_t^\star$; the net budgetary impact then reflects the other component of the subsidy, which counteracts monopsonistic price setting.

In conclusion, ceteris paribus a single-tier payment system requires lower taxes and generates fewer tax distortions than a two-tier system—even outside crisis periods. This can turn a preference for the two-tier system on the grounds of lower resource costs into a preference for the single-tier system once distortions are also accounted for. Such distortions may not only arise from the financing of subsidies but also from the deployment of nonfiscal instruments, for instance as a consequence of evasion, monitoring, or enforcement.

If a comparison of resource costs strongly favors a deposit-based system tax or similar distortions cannot rationalize a *circulating* CBDC. But they might still rationalize a *noncirculating* CBDC that disciplines banks. Recall that targeting a sufficiently high interest rate on CBDC (without issuing it in equilibrium) replicates the incentive effects of a deposit subsidy (proposition 4). When $\theta_t^\star$ requires more fiscal resources than the threat associated with a CBDC interest rate target and when taxes generate deadweight losses a noncirculating CBDC thus offers welfare gains. Summarizing:

Proposition 6. Suppose the social planner opts for deposits. Tax distortions provide a rationale for a circulating CBDC unless the two-tier system has substantially lower resource costs than a single-tier system. They provide a rationale for a noncirculating CBDC if $\theta_t^\star > 0$, independently of resource costs.

4.5 Central Bank Interest Rate Policy

A robust message of the first- and second-best analyses has been that money should pay interest and the interest rates on reserves and money, the two central bank liabilities, should differ.[55] Propositions 2 and 3 showed that the liquidity premium on a circulating CBDC should equal μ and that the premium on reserves, if they also circulate, should differ by $\omega_{1,t}(\zeta_{t+1}^\star, \zeta_{t+1}^\star) + \mu$. Intuitively, the optimal premium on reserves aims at implementing the first-best reserves-to-deposits ratio, which depends on the costs of liquidity substitution and externalities, while the optimum premium on CBDC only reflects payment operations costs.

When CBDC does not circulate but rather serves as a threat to discipline banks (with or without tax distortions), the optimal CBDC premium reflects the first-best deposit premium, $\nu + \omega_t(\zeta_{t+1}^\star, \zeta_{t+1}^\star) + \zeta_{t+1}^\star \rho$. Again, there is no reason to expect that this equals the optimum premium on reserves, $-\omega_{1,t}(\zeta_{t+1}^\star, \zeta_{t+1}^\star)$. Finally, in a second-best environment in which both CBDC and reserves circulate because the central bank targets ξ_{t+1} or m_{t+1} the spreads of the two instruments also typically differ. Summarizing:

Proposition 7. When the optimal policy involves both reserves and money the spreads on both liabilities generically optimally differ.

[55] Keister and Sanches (2022) analyze the optimality of differentiating interest rates on "cash-like" and "deposit-like" CBDC.

As noted before implementing the optimal policy therefore requires the government to price discriminate between wholesale and retail users of central bank liabilities.

5 Extensions

We finally review several other arguments made in CBDC discussions.

Price Rigidity, Effective Lower Bound Price rigidity in the production sector undermines the model's classical dichotomy and empowers monetary policy when producers are committed to supplying the quantities demanded (Yun, 1996; King and Wolman, 1996; Goodfriend and King, 1997). This holds independently of whether households use CBDC or deposits as means of payment and how monetary policy manages liquidity premia. Price rigidity does not structurally change the behavior of banks, which lies at the heart of our analysis.[56] Against this background we expect that price rigidity would add transmission channels rather than substantially altering the ones we have analyzed.

As discussed previously the presence of cash gives rise to an effective lower bound on nominal interest rates and this could be relevant if prices were rigid (Bordo and Levin, 2017). However, as pointed out by Buiter (2009) and others the presence of cash is conceptually distinct from an effective lower bound. More importantly, the introduction of CBDC is conceptually distinct from the abolition of cash.

Stimulus Payments Some observers argue that CBDC based stimulus payments would face fewer logistical difficulties than the payments made for instance in the context of the CARES Act of March 2020. But those difficulties reflected problems with address and death records, which a CBDC would also depend on. In the model, a "stimulus payment" in the form of a negative τ_t has the same effect as a helicopter CBDC drop, except that the latter changes real balances in addition to government transfers.

Bank Lending Frictions Absent lending frictions the model does not feature a bank lending channel (Bernanke and Blinder, 1988). To introduce such a channel we could assume that households cannot invest in (loans that fund) physical capital. Depending on the posited loan market structure this would introduce additional frictions. But unless these frictions interacted with the composition of bank funding they would likely not affect the mechanisms that drive our results; for example, they would not undermine the equivalence result.[57]

Frictions due to incentive or similar constraints could give rise to intermediary asset pricing (e.g., He and Krishnamurthy, 2018). Whether this would affect the equivalence

[56] For given inflation (expectations) the programs of banks that set nominal vs. real deposit rates are structurally equivalent. Of course, inflation affects money demand and price level determinacy requires a nominal anchor, e.g., a central bank interest rate rule satisfying the Taylor principle (Taylor, 1999; Woodford, 2001; Bullard and Mitra, 2002).

[57] See also the discussion in Keister and Sanches (2022, 6).

result or the normative considerations emphasized in the paper would depend on the extent the frictions interacted with the composition of bank funding.

Monetary Policy Targeting Some central banks consider caps on CBDC balances to limit the maximum size of deposit-CBDC transfers ("digital bank runs"). In an environment with heterogeneous households such caps could allow to target monetary policy if CBDC paid a higher interest rate than deposits and "poor" savers only held CBDC balances below the cap while "rich" savers held CBDC up to the cap as well as deposits. In such a situation the interest rate on CBDC would only have a direct substitution effect on the consumption and liquidity choices of poor savers.

6 Conclusion

One of the key questions triggered by the advent of CBDC is whether a single-tier (or mixed) monetary architecture is preferable to the conventional two-tier system. CBDC allows central banks to shortcut liquidity provision but this affects the role of banks. Our equivalence result suggests that the macro consequences may be manageable and, depending on central bank actions, quite limited unless the resource costs of single- and two-tier systems differ, deposits are "special" in ways that a central bank loan is not, or central bank balance sheet length is costly.

Whether central banks would *want* to insulate the financial sector and the macroeconomy from the consequences of CBDC is a different question. Central banks could choose to raise banks' funding costs—according to our funding-cost-reduction-at-risk measure by up to 1.5 percent of GDP or more—but they might also be able to improve on the equilibrium allocation in the two-tier system.

The optimal monetary arrangement provides liquidity at the lowest possible costs. In a CBDC based system the Ramsey policy follows the Friedman (1969) rule and charges a liquidity premium that reflects the payment operations costs of CBDC. In a deposit based system the optimal interest rate on reserves induces banks to internalize externalities from reserve holdings and a second instrument induces banks to pay interest on deposits that reflects the social costs.

The threat to introduce CBDC can help discipline banks independently of whether the single- or two-tier system provides liquidity at lowest resource costs. This is helpful when the government lacks instruments to address market power in liquidity provision as long as optimal reserve subsidies are not too high. It is also helpful when tax distortions or excess burdens from regulation render conventional instruments to discipline banks costly.

Frictions such as those underlying too-big-to-fail banks or tax distortions further strengthen the case for a circulating CBDC because they differentially raise the social costs of a two-tier system: Too-big-to-fail banks raise the effective resource costs of deposit based liquidity provision, and tax distortions render the single-tier system more attractive because providing incentives to banks in the two-tier system requires fiscal resources.

Whether the equilibrium allocation is first or second best and independently of whether

CBDC circulates or only serves to discipline banks, there is a strong case for differentiating the interest rates on reserves and CBDC because payment operations costs and externalities differ across means of payment. An important policy question therefore will be whether central banks can price discriminate between wholesale and retail users of central bank liabilities once they broaden access to central bank balance sheets.

Quantitatively, our analysis suggests cost advantages of the two-tier system in the baseline model. But these advantages are minor and easily overturned in the extensions with too-big-to-fail banks or tax distortions, implying that there is a case for a circulating CBDC, in addition to the case for a noncirculating CBDC to discipline banks.

Our workhorse macroeconomic model lends itself to many useful extensions, covering for instance open economy aspects or conflicting interests and political economy frictions, in addition to the extensions sketched in section 5. To sharpen the quantitative normative implications it could also be useful to model nonlinear substitution of money and deposits in households' effective real balances.

A Micro Foundations For Costs of Liquidity Substitution

Along the lines of Stein (2012) consider a bank that issues deposits, n_{t+1}, and equity and invests in capital, k_{t+1}. Stein (2012) assumes that deposits, which carry a liquidity premium (generating seignorage), can only be issued as long as their return is safe. As a consequence the price of capital in the worst state and the quantity of bank equity cap deposit issuance. Stein (2012) also assumes that the price of capital in the worst state is a decreasing function of aggregate capital investment, due to fire sales: To repay deposits a bank may sell capital to outside buyers but since these buyers operate a technology with decreasing returns their willingness to pay collapses when many banks sell (Shleifer and Vishny, 1992).

Stein (2012) shows that a bank confronted with fire sale risk has an incentive to increase the scale of its operations. But this imposes a negative pecuniary externality on other banks as more capital holdings lower the fire-sale price of capital, which constrains the activities of other banks and creates deadweight losses.[58] To correct the externality the central bank may cap money creation. In an extension Stein (2012) considers reserves, a minimum reserves requirement, and low interest on reserves as instruments to implement such a cap.

We modify this setup in three directions. First, we introduce reserves, r_{t+1}, from the beginning as a second type of bank asset. Second, we emphasize bank liquidity rather than solvency and interpret fire sales of capital as refinancing operations with the central bank (or on the interbank market), which occur at depressed prices, for instance because of asymmetric information about asset quality. Finally, we introduce a costly technology for the bank to make up for liquidity shortfalls. This cost is borne ex ante and reflects activities such as ex-ante information dissemination about the bank's portfolio or portfolio constraints along the lines discussed by King (2016).

Formally, let q_t denote the fire-sale price of capital. In Stein (2012) the fundamental bank constraint can be expressed as

$$q_t k_{t+1} \geq n_{t+1} \ldots,$$

indicating that deposit creation is constrained by worst case asset values. We replace this inequality by the condition that deposits must be covered by the sum of reserves, the lender-of-last-resort-support by the central bank, which equals $q_t k_{t+1}$, and the liquidity substitution afforded by the costly technology:

$$r_{t+1} + q_t k_{t+1} + \text{liquidity substitution} \geq n_{t+1}.$$

This inequality, which holds with equality if the bank minimizes costs and which can be written as

$$\text{liquidity substitution} = k_{t+1}(1 - q_t),$$

[58] Further social losses may arise when potential lenders freeze lending because they anticipate lucrative opportunities to buy at fire-sale prices (Diamond and Rajan, 2011).

implies that both a bank's individual reserve holdings and the aggregate reserve holdings decrease the bank's need for costly liquidity substitution. The former effect is present because reserve holdings lower a bank's capital holdings and the latter because aggregate reserve holdings depress q_t.

Let $\omega_t^{1/\varphi}$ denote the liquidity substitution effect, per unit of deposit, where ω_t denotes the unit cost of generating this effect. Decreasing returns imply $\varphi > 1$. Furthermore, let the price of capital be given by

$$q_t = 1 - \phi_t^{1/\varphi}(1 - \bar{\zeta}_{t+1})^{\bar{\varphi}/\varphi}, \quad \phi_t > 0, \ \bar{\varphi}/\varphi > 0;$$

that is, there is no negative price impact when $\bar{\zeta}_{t+1} = 1$; the minimum price $1-\phi_t^{1/\varphi}$ results when $\bar{\zeta}_{t+1} = 0$; and q_t is a decreasing function of aggregate capital exposure. Using the definition of ζ_{t+1} we obtain

$$\frac{\text{liquidity substitution}}{n_{t+1}} = \omega_t^{1/\varphi} = (1 - \zeta_{t+1})\phi_t^{1/\varphi}(1 - \bar{\zeta}_{t+1})^{\bar{\varphi}/\varphi}$$

or

$$\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1}) = (1 - \zeta_{t+1})^{\varphi}\phi_t(1 - \bar{\zeta}_{t+1})^{\bar{\varphi}}.$$

This specification satisfies the conditions on function ω_t stated in the text.

B General Equilibrium Under Assumptions 1 and 2

B.1 Households Under Assumption 1

Let $\mathcal{A}_t \equiv (1 - \vartheta)c_t^{1-\psi} + \vartheta z_{t+1}^{1-\psi}$ such that the marginal utility of consumption and real balances, respectively, is given by

$$\begin{aligned} u_c(c_t, z_{t+1}, x_t) &= (1 - \vartheta)\mathcal{A}_t^{\frac{1-\sigma}{1-\psi}-1} c_t^{-\psi}, \\ u_z(c_t, z_{t+1}, x_t) &= \vartheta\mathcal{A}_t^{\frac{1-\sigma}{1-\psi}-1} z_{t+1}^{-\psi}. \end{aligned}$$

When deposits circulate, $n_{t+1} > 0$ and possibly $m_{t+1} > 0$ as well, the Euler equation for real balances, equation (4), reduces to

$$\frac{\vartheta z_{t+1}^{-\psi}}{(1 - \vartheta)c_t^{-\psi}} = \chi_{t+1}^n \quad \text{or} \quad z_{t+1} = c_t\left(\frac{\vartheta}{1 - \vartheta}\frac{1}{\chi_{t+1}^n}\right)^{\frac{1}{\psi}}. \tag{4'}$$

When deposits do not circulate, $n_{t+1} = 0$ and $m_{t+1} > 0$, equation (4) reduces to

$$\lambda_t\frac{\vartheta(\lambda_t m_{t+1})^{-\psi}}{(1 - \vartheta)c_t^{-\psi}} = \chi_{t+1}^m \quad \text{or} \quad z_{t+1} = \lambda_t m_{t+1} = c_t\left(\frac{\vartheta}{1 - \vartheta}\frac{\lambda_t}{\chi_{t+1}^m}\right)^{\frac{1}{\psi}}. \tag{4''}$$

Let χ_{t+1} denote "the" spread on household means of payment. By convention, $\chi_{t+1} = \chi_{t+1}^m/\lambda_t$ when money circulates; $\chi_{t+1} = \chi_{t+1}^n$ when deposits circulate; and $\chi_{t+1} = \chi_{t+1}^m/\lambda_t =$

χ^n_{t+1} when both money and deposits circulate (see equation (5)). Moreover, define

$$\Omega^c_t \equiv (1-\vartheta)^{\frac{1-\sigma}{1-\psi}} \left(1 + \left(\frac{\vartheta}{1-\vartheta}\right)^{\frac{1}{\psi}} \chi_{t+1}^{1-\frac{1}{\psi}}\right)^{\frac{\psi-\sigma}{1-\psi}}.$$

We can then express the marginal utility of consumption conditional on equilibrium real balances as $c_t^{-\sigma}\Omega^c_t$, independently of whether money, deposits, or both means of payment circulate. The household's optimality conditions conditional on equilibrium real balances thus read

$$c_t^{-\sigma}\Omega^c_t = \beta\mathbb{E}_t\left[R^k_{t+1}c^{-\sigma}_{t+1}\Omega^c_{t+1}(\chi_{t+2})\right] \tag{2'}$$

for capital and

$$v'(x_t) = c_t^{-\sigma}\Omega^c_t w_t \tag{3'}$$

for leisure.[59]

B.2 Resource Constraint Under Assumption 1

Recall that $\xi_{t+1} \equiv n_{t+1}/z_{t+1}$ denotes the share of deposits in effective real balances. Defining

$$\Omega^{rc}_t \equiv 1 + \left(\frac{\vartheta}{1-\vartheta}\frac{1}{\chi_{t+1}}\right)^{\frac{1}{\psi}} \left(\xi_{t+1}\left\{\nu + \omega_t + \zeta_{t+1}\rho\right\} + (1-\xi_{t+1})\frac{\mu}{\lambda_t}\right),$$

we can express the resource constraint subject to equilibrium real balances as

$$\kappa_{t+1} = f_t(\kappa_t, 1-x_t) + \kappa_t(1-\delta) - c_t\Omega^{rc}_t, \tag{15'}$$

where we use equations (4') and (4").

B.3 Banks

B.3.1 Deposit Spread Under Assumption 1

The spread on deposits depends on monetary policy. We consider several cases:

Central Bank Targets Composition of Real Balances When the central bank targets $\xi_{t+1} \equiv n_{t+1}/z_{t+1}$ equation (4') yields

$$n_{t+1} = \xi_{t+1}c_t\left(\frac{\vartheta}{1-\vartheta}\frac{1}{\chi^n_{t+1}}\right)^{\frac{1}{\psi}} \quad \Rightarrow \quad \eta_{n,t+1} \equiv \frac{{n_{t+1}}'(R^n_{t+1})R^n_{t+1}}{n_{t+1}(R^n_{t+1})} = \frac{1}{\psi}\frac{R^n_{t+1}}{R^f_{t+1}-R^n_{t+1}}$$

and the bank's optimality condition for deposits, equation (8a), reduces to

$$\chi^n_{t+1} = \frac{\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t}{1-\psi}. \tag{8a'}$$

[59]With a multiplicatively rather than additively separable specification of preferences the household first-order conditions also feature a pseudo wedge $\Omega^x_t(\chi_{t+1})$ multiplying $v'(x_t)$; see Niepelt (2020a).

Central Bank Targets Quantity of Money In this case, $n_{t+1} = z_{t+1} - \lambda_t m_{t+1}$ and ξ_{t+1} is endogenous. Equation (4') yields

$$n_{t+1} = c_t \left(\frac{\vartheta}{1-\vartheta} \frac{1}{\chi^n_{t+1}} \right)^{\frac{1}{\psi}} - \lambda_t m_{t+1} \quad \Rightarrow \quad \eta_{n,t+1} = \frac{1}{\psi} \frac{R^n_{t+1}}{R^f_{t+1} - R^n_{t+1}} \frac{1}{\xi_{t+1}}.$$

Accordingly, the bank's optimality condition reduces to

$$\chi^n_{t+1} = \frac{\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t}{1 - \psi \xi_{t+1}}. \tag{8a''}$$

When the central bank does not issue money, $\xi_{t+1} = 1$ and (8a') and (8a") coincide.

Central Bank Targets Spread on Money We refer to the profit maximizing deposit liquidity premium χ^n_{t+1} in equation (8a') as the "monopsony premium." When the central bank targets χ^m_{t+1} and sets it in excess of λ_t times the monopsony premium (rendering money unattractive for households) then monetary policy is irrelevant and $m_{t+1} = 0$.

When the central bank targets χ^m_{t+1} and sets it below λ_t times the monopsony premium (rendering money attractive) then banks have a choice between raising the interest rate on deposits to the competitive rate given in equation (5) (plus epsilon), thereby pricing the central bank out of the market, or not raising the rate and being priced out of the market themselves. As long as the deposit premium covers costs,

$$\chi^n_{t+1} \geq \nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t + \zeta_{t+1}\chi^r_{t+1} = \nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t,$$

banks optimally choose the former option, that is, they satisfy equation (5).

In conclusion, when the central bank targets χ^m_{t+1} there are three cases to distinguish:

(a) When $\frac{\chi^m_{t+1}}{\lambda_t} < \nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t$ banks would lose money when trying to compete with the central bank. In equilibrium, $n_{t+1} = 0$.

(b) When $\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t \leq \frac{\chi^m_{t+1}}{\lambda_t} < \frac{\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t}{1-\psi}$ banks issue deposits at a premium lower than the monopsony premium and equal to (slightly less than) χ^m_{t+1}/λ_t in order to compete with the central bank. In equilibrium, $m_{t+1} = 0$.

(c) When $\frac{\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t}{1-\psi} \leq \frac{\chi^m_{t+1}}{\lambda_t}$ banks issue deposits at the monopsony premium; monetary policy is irrelevant. In equilibrium, $m_{t+1} = 0$.

B.3.2 Reserve Holdings and Liquidity Substitution Costs Under Assumption 2

The bank's optimality condition for reserves, equation (9), implies

$$\phi_t \varphi (1 - \zeta_{t+1})^{\varphi - 1} (1 - \bar{\zeta}_{t+1})^{\bar{\varphi}} = \chi^r_{t+1}$$

such that in equilibrium

$$\zeta_{t+1} = 1 - \left(\frac{\chi^r_{t+1}}{\phi_t \varphi} \right)^{\frac{1}{\varphi + \bar{\varphi} - 1}}. \tag{9'}$$

(We only consider $\zeta_{t+1} \in (0,1)$.) Accordingly, the equilibrium costs of liquidity substitution equal

$$\omega_t(\zeta_{t+1}, \zeta_{t+1}) = \phi_t \left(\frac{\chi^r_{t+1}}{\phi_t \varphi}\right)^{\frac{\varphi+\bar{\varphi}}{\varphi+\bar{\varphi}-1}} \tag{16}$$

and, using the fact that $\tilde{\omega}_t(-\chi^r_{t+1}) = \omega_t(\zeta_{t+1}, \zeta_{t+1}) + \chi^r_{t+1}\zeta_{t+1}$ (see the main text after equation (8a)),

$$\tilde{\omega}_t(-\chi^r_{t+1}) = \chi^r_{t+1} - (\varphi - 1)\phi_t \left(\frac{\chi^r_{t+1}}{\phi_t \varphi}\right)^{\frac{\varphi+\bar{\varphi}}{\varphi+\bar{\varphi}-1}}. \tag{17}$$

B.4 General Equilibrium Under Assumptions 1 and 2

Combining these results we conclude that an equilibrium allocation, price, and payment system satisfies conditions (11), (12), (2'), (3'), (15') as well as the following restrictions:

- In a monetary system with deposits, $\chi_{t+1} = \chi^n_{t+1}$ as well as conditions (4'), (9'), (16), (17) and
  - when the central bank targets ξ_{t+1}, condition (8a') and $m_{t+1} = (1-\xi_{t+1})z_{t+1}/\lambda_t$;
  - when it targets m_{t+1}, condition (8a") and $\xi_{t+1} = 1 - \lambda_t m_{t+1}/z_{t+1}$;
  - when it targets χ^m_{t+1}, the conditions described under (b) or (c) above.
- In a monetary system without deposits, $\chi_{t+1} = \chi^m_{t+1}/\lambda_t$ as well as condition (4").

C Calibration

We calibrate the model to conform with a quarterly frequency. We posit a Cobb-Douglas production function with capital share α and following standard practice (e.g. Gertler and Karadi, 2011) we let $\alpha = 0.33$; $\beta = 0.99$; $\delta = 0.025$; $\sigma \to 1$; and $v(x) = -3.409(1-x)^{1.276}/1.276$, implying an inverse Frisch elasticity of 0.276.

FDIC historical data since 2010 imply average annual bank net interest margins of 0.0323 and average annual returns on assets of 0.0102.[60] This corresponds to a quarterly deposit liquidity premium $\chi = 0.0323/4/1.005$ and quarterly costs relative to assets of $(0.0323 - 0.0102)/4$.[61] Drechsler et al. (2021) confirm that bank margins have been very stable over time. Since 2010 the annual interest rate on reserves exceeded the annual interest rate on deposits (including money market deposit accounts) by on average $0.0068 - 0.0007$, and the reserves-to-deposits ratio averaged roughly 0.25.[62] A substantial

[60] See *Statistics at a Glance*, retrieved from FDIC; https://www.fdic.gov/analysis/quarterly-banking-profile/statistics-at-a-glance/2021dec/fdic.xlsx.

[61] An annual liquidity premium, $\chi^i = 1 - R^i/R^f$, corresponds to a quarterly premium $1 - \sqrt[4]{1-\chi^i}$. For χ^i close to zero this is approximately equal to $\chi^i/4$. When converting interest rate differentials into liquidity premia we assume that the annual nominal risk-free rate equals two percent.

[62] See appendix G for data sources.

share of measured bank costs are fixed costs.[63] Since fixed costs do not enter the bank's first-order condition (8a') for the deposit spread we scale the measured costs by a factor of 0.5 to infer marginal costs. Finally, we use the fact that $\tilde{\omega} = \omega + \chi^r\zeta$. This yields the following conditions:

$$\begin{aligned}
\chi &= 0.0323/4/1.005, \\
\chi^r &= \chi - (0.0068 - 0.0007)/4/1.005, \\
\zeta &= 0.25, \\
\frac{(\nu + \omega - \theta)n}{n/(1-\zeta)} = (\nu + \omega - \theta)(1-\zeta) &= (0.0323 - 0.0102)/4 \cdot 0.5, \\
\frac{\nu + \omega + \chi^r\zeta - \theta}{1-\psi} &= \chi.
\end{aligned}$$

Di Tella (2020, p. 2002) computes a liquidity share in expenditures of 0.017. Using the household first-order condition (4') in appendix B thus implies

$$\left(\frac{z\chi}{c}\right)^{\psi} = \frac{\vartheta}{1-\vartheta}\chi^{\psi-1}, \quad \frac{z\chi}{c} = 0.017.$$

Schmiedel et al. (2012) report costs for banks and central banks of managing retail payments in the Euro area on the order of 0.5 and 0.03 percent of GDP, respectively.[64] We posit the same (relative) magnitudes for the US, and we use the fact that in the U.S. the ratio of deposits to annual GDP equalled roughly eighty percent during the last decade.[65] We assume that the costs of managing money-based payments would (in the most likely scenario in which banks interact with retail customers) be equal to the costs of managing deposit-based payments in a narrow banking system, $\mu = \nu + 1\rho$. This yields the following conditions:

$$\begin{aligned}
\frac{\rho\zeta}{\nu} &= \frac{0.03}{0.5}, \\
\nu\ 0.8 &= 0.005/4, \\
\mu &= \nu + \rho.
\end{aligned}$$

Shleifer and Vishny (2011) review the evidence on fire sale prices of securities and real assets. We set the maximum price impact $\phi^{1/\varphi} = 0.2$ (see appendix A). Together with the definition of ω and the bank equilibrium condition for reserve holdings, equation (9') in appendix B, this implies

$$\begin{aligned}
\phi^{1/\varphi} &= 0.2, \\
\omega &= \phi(1-\zeta)^{\varphi+\bar{\varphi}}, \\
\zeta &= 1 - \left(\frac{\chi^r}{\phi\varphi}\right)^{\frac{1}{\varphi+\bar{\varphi}-1}}.
\end{aligned}$$

[63] See *Bank's Overhead Costs to Total Assets for United States* [DDEI04USA156NWDB], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DDEI04USA156NWDB.

[64] Costs for retailers equal 0.46 percent of GDP. Cash and debit card transactions exhibit similar costs per transaction.

[65] See *Bank Deposits to GDP for United States* [DDOI02USA156NWDB], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DDOI02USA156NWDB.

The above system of equations can be solved for $\psi, \vartheta, \mu, \nu, \rho, \omega, \phi, \bar{\varphi}, \theta$ once we specify a value for φ. We choose the value for φ that implies $\theta = 0$; different choices change the calibration of ω and θ but not any other parameters.

Table 1 summarizes the calibration. The implied liquidity substitution costs are $\omega = 0.0021$ and the implied capital-to-deposits ratio equals 5.8551, which compares with roughly $4/0.8 = 5$ in the data. We let $\lambda = 1$.

Parameter	Value	Source
Production		
α	0.3300	Gertler and Karadi (2011)
δ	0.0250	Gertler and Karadi (2011)
Preferences		
β	0.9900	Gertler and Karadi (2011)
σ	1.0000	Gertler and Karadi (2011)
ψ	0.3388	FDIC, FRED, model
ϑ	0.0103	Di Tella (2020), model
$v(x)$	$-\frac{3.4090}{1.2760}(1-x)^{1.2760}$	Gertler and Karadi (2011)
Payments and liquidity substitution		
μ	0.0019	Schmiedel et al. (2012), FRED, model
ν	0.0016	Schmiedel et al. (2012), FRED
ρ	0.0004	Schmiedel et al. (2012), FRED
ϕ	0.0245	Shleifer and Vishny (2011), FRED, model
φ	2.3048	Shleifer and Vishny (2011), FRED, model
$\bar{\varphi}$	6.1995	Shleifer and Vishny (2011), FRED, model
θ	0.0000	Assumption
λ	1.0000	Assumption

Table 1: Calibration for economy with deposit-based payments

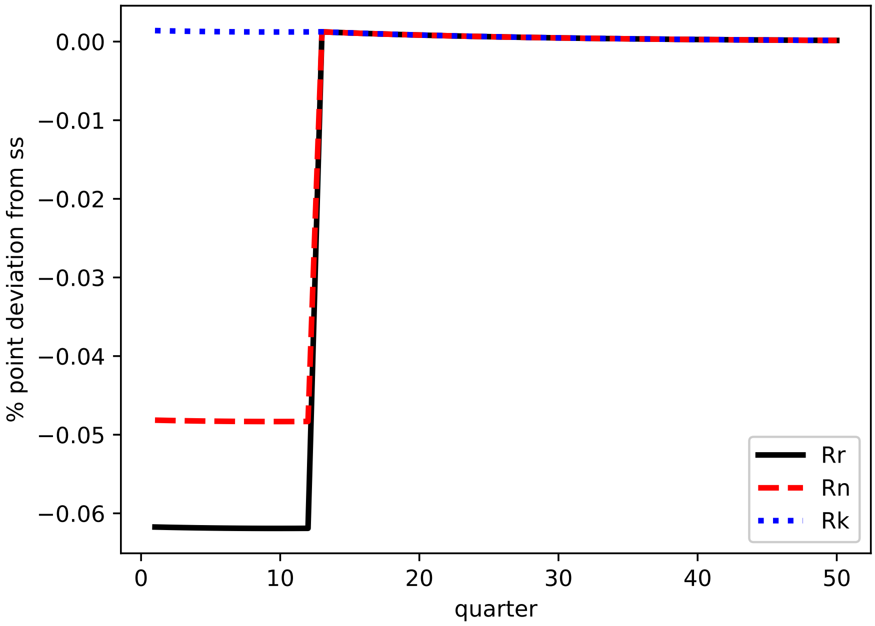
D Impulse Response Functions

D.1 Increase of χ^r_{t+1}

Figure 1 in the main text and figure 3 below illustrate the effects of an increase in χ^r_{t+1} by 25 basis points (on an annual basis) for twelve quarters.

D.2 Increase of θ_t

Figure 4 below illustrates the effects of an increase in θ_t by 25 basis points (on an annual basis) for twelve quarters. That change increases R^n_{t+1} by roughly nine basis points and r_{t+1} and n_{t+1} by nearly 35 percent.



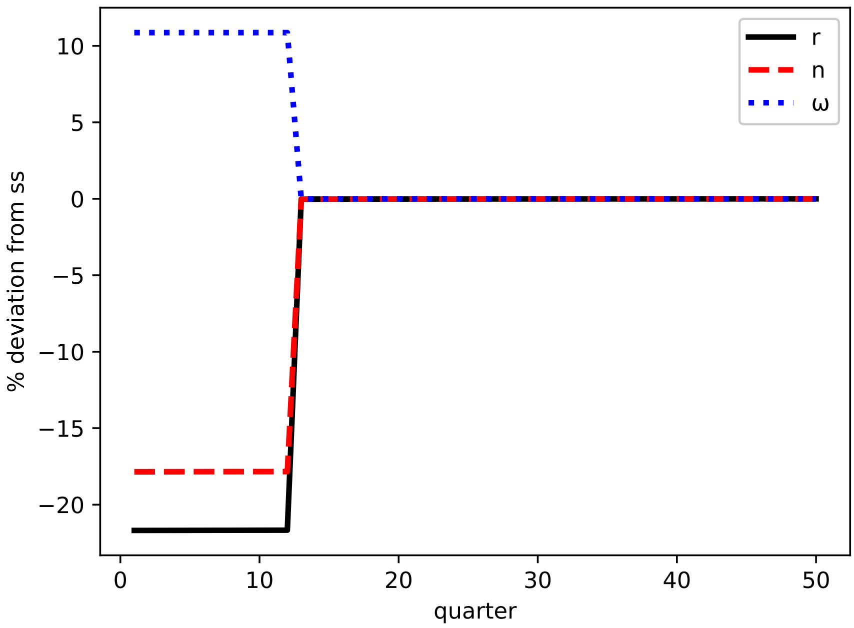


Figure 3: Responses to an increase of χ^r_{t+1} by 25 basis points (annual) at $t = 1$ that lasts for twelve quarters.

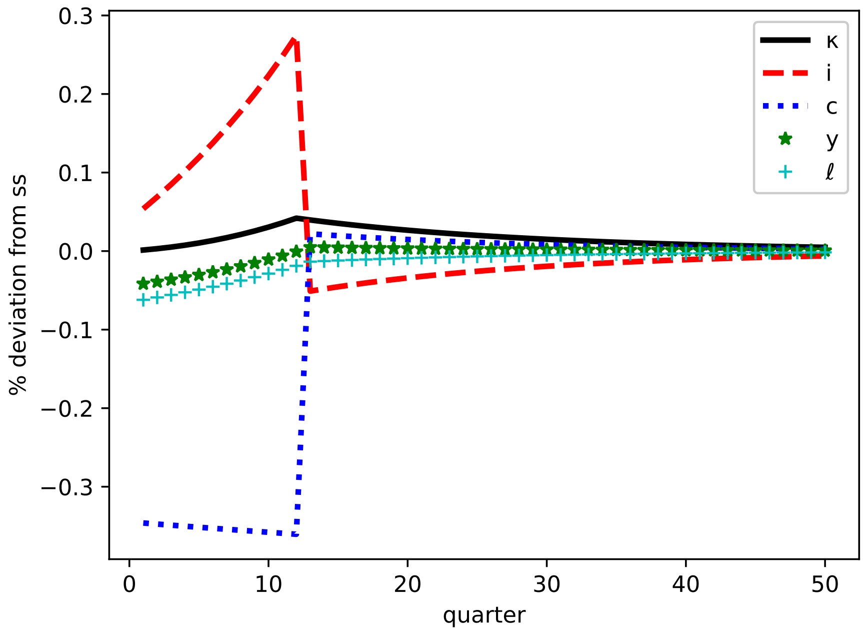


Figure 4: Responses to an increase of θ_t by 25 basis points (annual) at $t = 1$ that lasts for twelve quarters.

E Equivalence

We state proposition 1 formally, derive its implications for capital accumulation, portfolios, and budget sets, and prove it before discussing its implications and robustness. We consider an intervention which reduces deposit holdings at date t by $\Delta > 0$ and increases money holdings at date t by $\lambda_t^{-1}\Delta$. To guarantee nonnegativity of deposits, capital holdings, and reserves the intervention Δ must not be too large.[66] For generality we allow the returns on money and deposits to be stochastic.

Proposition Suppose condition 1 holds. Consider a policy that implements an equilibrium with deposits and reserves. There exists another policy and equilibrium, indicated by circumflexes, with fewer deposits and reserves, more money, a central bank loan, a different ownership structure of capital, possibly household taxes at dates t and $t+1$ whose market value equals zero, and otherwise the same allocation and price system.[67]

[66]Specifically, we impose $\Delta \leq n_{t+1}$, $(1-\lambda_t^{-1})\Delta \leq k^g_{t+1}$, $(1-\lambda_t^{-1})\Delta \geq -k_{t+1}$, and $\zeta_{t+1}\Delta \leq r_{t+1}$. The assumption that the central bank loan in the initial equilibrium equals zero amounts to a convenient normalization.

[67]If the initial policy does not include CBDC then the introduction of CBDC must not change the asset span.

The two equilibria coincide except that

$$\hat{m}_{t+1} = m_{t+1} + \lambda_t^{-1}\Delta, \quad \hat{n}_{t+1} = n_{t+1} - \Delta, \quad \hat{l}_{t+1} = \Delta(1-\zeta_{t+1}), \quad \hat{r}_{t+1} = r_{t+1} - \zeta_{t+1}\Delta,$$
$$\hat{k}_{t+1} = k_{t+1} + (1-\lambda_t^{-1})\Delta, \quad \hat{k}^g_{t+1} = k^g_{t+1} - (1-\lambda_t^{-1})\Delta.$$

The household tax at date t, $\hat{T}_{1,t}$, and the state contingent tax at date $t+1$, $\hat{T}_{2,t+1}$, satisfy

$$\begin{aligned} \hat{T}_{1,t} &= \Delta\{\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t\}, \\ \hat{T}_{2,t+1} &= \Delta\{(1-\lambda_t^{-1})R^k_{t+1} + \lambda_t^{-1}R^m_{t+1} - R^n_{t+1} - (\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t)R^f_{t+1}\}. \end{aligned}$$

The central bank loan carries the interest rate

$$R^l_{t+1} = \frac{R^n_{t+1} + (\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t)R^f_{t+1} - \zeta_{t+1}R^r_{t+1}}{1-\zeta_{t+1}}.$$

Implications for Capital Accumulation, Portfolios, and Budget Sets The portfolio and policy changes described in the proposition have several implications. First, $\hat{z}_{t+1} = z_{t+1}$, $\hat{m}_{t+1}\mu + \hat{n}_{t+1}(\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1})) + \hat{r}_{t+1}\rho = m_{t+1}\mu + n_{t+1}(\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1})) + r_{t+1}\rho$, and $\hat{\kappa}_{t+1} = \kappa_{t+1}$; that is, the portfolio changes do not alter effective real balances, the aggregate capital stock, or the total resource costs of operating the payment system and transforming maturity. Note that the portfolio changes also leave the reserves-to-deposits ratio, ζ_{t+1}, and thus $\omega_t(\zeta_{t+1}, \zeta_{t+1})$ unchanged.

Second, the length of household balance sheets does not change since households swap assets (deposits, money, and capital). In contrast, the balance sheets of banks shorten because banks hold fewer reserves and reduce borrowing (deposits and central bank loans) but invest the same amount in capital as before the intervention. The balance sheet of the consolidated government expands by $(\lambda_t^{-1} - \zeta_{t+1})\Delta \lesseqgtr 0$: The central bank raises additional funds Δ/λ_t from households but fewer funds from banks, $-\zeta_{t+1}\Delta$; it passes $\Delta(1-\zeta_{t+1})$ through to the banking sector and increases capital holdings by $(\lambda_t^{-1} - 1)\Delta$.

Third, the new tax at date t compensates for the reduced equity purchases by households, $\hat{\pi}^b_{1,t} - \pi^b_{1,t} = \hat{T}_{1,t}$, and the new state contingent taxes at date $t+1$ compensate for the change in the return on the household portfolio at date $t+1$. Formally, we have $(\hat{k}_{t+1} - k_{t+1})R^k_{t+1} + (\hat{m}_{t+1} - m_{t+1})R^m_{t+1} + (\hat{n}_{t+1} - n_{t+1})R^n_{t+1} = (1-\lambda_t^{-1})\Delta R^k_{t+1} + \lambda_t^{-1}\Delta R^m_{t+1} - \Delta R^n_{t+1}$ and $\hat{\pi}^b_{2,t+1} - \pi^b_{2,t+1} = -\Delta(R^k_{t+1} - R^n_{t+1}) + \Delta(1-\zeta_{t+1})(R^k_{t+1} - R^l_{t+1}) - \zeta_{t+1}\Delta(R^r_{t+1} - R^k_{t+1}) = -\Delta(\nu + \omega_t(\cdot) - \theta_t)R^f_{t+1}$, where we use the expression for the loan rate. The market value of the two taxes as of date t equals zero, $\hat{T}_{1,t} + \mathbb{E}_t[\text{sdf}_{t+1}\hat{T}_{2,t+1}] = 0$, and so does the market value of the changes in bank cash flows. Formally, the market value of the two taxes equals Δ times $(\nu + \omega_t(\cdot) - \theta_t) + (1-\lambda_t^{-1}) + \lambda_t^{-1}\mathbb{E}_t[\text{sdf}_{t+1}R^m_{t+1}] - \mathbb{E}_t[\text{sdf}_{t+1}R^n_{t+1}] - (\nu + \omega_t(\cdot) - \theta_t) = 0$,[68] and the market value of the change in bank cash flows equals $\hat{\pi}^b_{1,t} - \pi^b_{1,t} + \mathbb{E}_t[\text{sdf}_{t+1}(\hat{\pi}^b_{2,t+1} - \pi^b_{2,t+1})] = \hat{T}_{1,t} - \Delta(\nu + \omega_t(\cdot) - \theta_t) = 0$. As a consequence, the household's dynamic and intertemporal budget constraints continue to be satisfied with the modified portfolios and policy.

[68] When the interest rates on money and deposits are stochastic, indifference between the two means of payment does not imply equation (5). Instead, the Euler equations for m_{t+1} and n_{t+1} imply $\mathbb{E}_t[\text{sdf}_{t+1}(\lambda_t^{-1}R^m_{t+1} - R^n_{t+1})] = \lambda_t^{-1} - 1$. The result then follows.

Fourth, the same holds true for the government. From equation (13), the government budget constraint at date t reads $\hat{k}^g_{t+1} + \hat{l}_{t+1} - \hat{m}_{t+1} - \hat{r}_{t+1} = k^g_t R^k_t - m_t R^m_t - r_t R^r_t + \tau_t - \hat{n}_{t+1}\theta_t - \hat{m}_{t+1}\mu - \hat{r}_{t+1}\rho + \hat{T}_{1,t}$. Condition 1 implies that this equality is satisfied if and only if the budget constraint was satisfied before the intervention. Similarly, the government budget constraint at date $t+1$, $k^g_{t+2} - m_{t+2} - r_{t+2} = \hat{k}^g_{t+1} R^k_{t+1} + \hat{l}_{t+1} R^l_{t+1} - \hat{m}_{t+1} R^m_{t+1} - \hat{r}_{t+1} R^r_{t+1} + \tau_{t+1} - n_{t+2}\theta_{t+1} - m_{t+2}\mu - r_{t+2}\rho + \hat{T}_{2,t+1}$, is equivalent to the constraint before the change of portfolios and policy.

Proof Conjecture that the price system does not change, as claimed in the proposition. The optimal production decisions of firms are unchanged in this case, as are firm profits. Moreover, the market values of households' time endowments, taxes, and bank cash flows (as shown above) also do not change. As a consequence, household wealth is unaffected, and so is demand for consumption, leisure, and real balances. As shown above, this demand is supported by the modified portfolios. It remains to be shown that the modified bank portfolios are optimal; in that case, all budget constraints are satisfied at the optimal choices, equilibrium capital accumulation is unchanged, and the conjecture is verified.

To render deposits $\hat{n}_{t+1}$, loans $\hat{l}_{t+1}$, and reserves $\hat{r}_{t+1}$ optimal for a bank it suffices for the central bank to structure the loans such that the bank's choice sets before and after the intervention coincide. Before the intervention, this choice set is determined by the cost function, ω_t; the subsidy, θ_t; the deposit funding schedule; the stochastic discount factor; and the returns on capital and reserves. After the intervention, it is defined by the same cost function, subsidy rate, stochastic discount factor, and returns; a modified deposit funding schedule (because households hold more money); and a central bank loan funding schedule.

To assure identical choice sets it therefore suffices for the central bank to post an appropriate loan funding schedule. This schedule makes up for the shift in the deposit funding schedule, corrected for the fact that $1 - \zeta_{t+1}$ dollars of central bank loans provide the same net funding as one dollar of deposits of which ζ_{t+1} dollars are invested in reserves.[69] Subject to the appropriate loan funding schedule a bank chooses loans that make up for the reduction in funding (net of reserves) from households, at the same effective price; moreover, it chooses the same reserves-to-deposits ratio such that $\omega_t(\zeta_{t+1}, \zeta_{t+1})$ is unaffected.

Discussion Several remarks on the interpretation of proposition 1 as well as its robustness and possible sources of nonequivalence are in order. Consider first the interpretation:

i. The equivalence result states that a given equilibrium allocation and price system is associated with multiple compositions of real balances cum policies. This multiplicity is distinct from the instrument redundancy discussed in sections 2 and 4 according to which multiple policies may implement the same equilibrium including a specific portfolio structure.

[69] If the banking sector were competitive no pass through funding would be required; equivalence could also be achieved if banks shed assets, see Brunnermeier and Niepelt (2019).

ii. The equivalence characterized in the proposition has normative implications: Under condition 1 and starting from any policy and equilibrium the government *can* "sterilize" the equilibrium effects of an introduction or expansion of money. Under condition 1 welfare in an economy with deposits therefore constitutes a lower bound to welfare in an optimally managed economy with CBDC (and deposits). A strict welfare improvement could arise if the government did not act optimally in the initial equilibrium, or if the introduction of CBDC opens up options that were not available in a two-tier system without CBDC.[70]

iii. Empirically, many central banks today have substantially longer balance sheets than before the financial crisis, corresponding to high ζ_{t+1} values in the banking system. Since the volume of central bank loans under the equivalent policy is a decreasing function of the initial reserves-to-deposits ratio, an introduction of CBDC (under condition 1 and the equivalent policy) thus would require smaller central bank loans today than fifteen years ago.

The proposition follows under much more general conditions than those laid out in section 2:

i. The equivalence result does not hinge on specific preferences or types of monetary friction (money in the utility function or otherwise). The logic of the result is based on the stability of choice sets rather than specific first-order conditions.

ii. Nor does the equivalence result hinge on a specific market structure in the banking sector. While the market structure shapes the equivalent loan funding schedule that preserves the choice sets of banks, no conditions on that schedule are imposed.

iii. The equivalence result applies for arbitrary initial policies as long as they implement an equilibrium. For example, it applies without change when the initial policy employs only a subset of instruments (setting $\theta_t = 0$, say).

iv. The equivalence result also applies in the presence of bank runs since we allow for stochastic returns on deposits (and money). When deposits are risky but money is risk-free equivalence requires state-contingent transfers (with a market value of zero) to neutralize changes in the return characteristics of household portfolios.

v. Fixed costs of payment networks would not undermine the equivalence result as long as the same types of fixed costs were borne in the initial and the new equilibrium. This appears plausible for the realistic scenario in which banks rather than the central bank would be the ones to face customers making CBDC based payments.

vi. Household or bank heterogeneity, e.g., household specific λ_t parameters or bank specific ω_t functions, would not necessarily undermine the equivalence result either. If condition 1 were satisfied for a specific group but not for others the substitution

[70] For example, Brunnermeier and Niepelt (2019) argue that rather than replicating households' deposit holding strategies the central bank as a "large" lender could do better by internalizing bank run externalities.

of money for deposits would remain neutral as long as an appropriately targeted policy accompanied it.

Finally, the proof of proposition 1 points to possible causes of nonequivalence some of which the literature has touched upon:

i. Violation of condition 1 necessarily undermines equivalence. When the introduction or extension of CBDC changes resource requirements it is impossible to preserve the allocation and liquidity provision.

ii. If deposits and money cannot be linearly substituted as components of household effective real balances, i.e., if λ_t is endogenous to the composition of real balances, then "average λ_t" and "marginal λ_t" differ and the proof of proposition 1 no longer goes through. Similar arguments would apply for firms or banks if they hat to satisfy constraints (other than budget or balance sheet constraints) that depend on real balances. Brunnermeier and Niepelt (2019) refer to the condition that deposits and money can linearly be substituted as "liquidity neutrality."

iii. If deposits affect the choice sets of banks not only as a source of funding but also through other channels, and if central bank loans do not have these additional effects, then loans and deposits are imperfect substitutes and this undermines equivalence. Nonsubstitutability arises, for example, if deposits but not central bank loans are "special" because the former induce more monitoring. (This is unlikely given that banks are mainly monitored by the central bank and bank supervisors.)[71]

    Nonsubstitutability also arises when banks need to post less collateral for deposit funding than for central bank loans, and if collateral is scarce, as in Williamson (2019) or Böser and Gersbach (2020). Given that central banks typically only provide secured funding[72] one might conclude that an unsecured loan—and thus, equivalence—is "unrealistic." But the situation is more complex. Unsecured loans under the equivalent policy with CBDC are the mirror image of unsecured implicit lender-of-last-resort guarantees that provide the liquidity backing for deposits in the two-tier system. If those implicit guarantees were secured then the equivalent loans would be secured as well. Rather than prescribing an unrealistic policy the equivalence result points to a potential inconsistency of central bank policies.

iv. If gross balance sheet positions directly affect resource costs then passthrough funding affects the allocation and the equivalence result may not hold. Such a situation arises for example when central bank assets (and thus, passthrough funding) are costly to manage or when households and the central bank incur different costs

[71]Synergies between bank assets and liabilities such as in Kashyap et al. (2002) or Hanson et al. (2015) do not undermine equivalence as long as the synergies are also present with a central bank loan. Pulley and Humphrey (1993) offer an empirical assessment of such synergies.

[72]Collateral requirements reflect concerns about central bank net worth and independence. In the model the central bank need not guard against losses on its assets because the government sector is consolidated, has access to nondistorting taxes, and households are homogeneous.

when holding physical capital. Piazzesi and Schneider (2021) consider an environment in which it is costly to manage bank, central bank, but not household assets, and where only banks can offer contingent, on-demand liquidity via credit lines, which requires fewer assets to back than noncontingent deposits/CBDC.[73]

v. If the central bank lacks the information to post the equivalent loan supply schedule (reflecting the deposit funding schedule that would prevail without CBDC) then the equivalent policy cannot be implemented (Niepelt, 2020b). Nonequivalence also follows when the switch from deposits to CBDC affects learning and innovation.

vi. If the introduction of CBDC changes the asset span the equivalence result may not apply (Brunnermeier and Niepelt, 2019).[74] Similarly, it may not apply when the government cannot impose the taxes $\hat{T}_{1,t}$ or $\hat{T}_{2,t+1}$.

vii. If the central bank injects CBDC by transfer rather than absorbing deposits (and their liquidity services) in exchange, the market for real balances clears at a modified interest rate and the equilibrium allocation changes. Keister and Sanches (2022, 6) discuss a related effect in their framework. Also, if the central bank issues CBDC in exchange for government bonds, as in Kumhof and Noone (2021) and Barrdear and Kumhof (2022), it prevents the deposit-CBDC substitution characterized in the equivalence result.

F Bank Funding-Cost-Reduction-at-Risk

Recall that R^l_{t+1}, which can be expressed as

$$R^r_{t+1} + \frac{R^n_{t+1} - R^r_{t+1}}{1 - \zeta_{t+1}} + \frac{(\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t) R^f_{t+1}}{1 - \zeta_{t+1}}, \tag{18}$$

represents the costs for a bank in the two-tier system to finance one unit of capital investment with deposits. If the bank borrowed from the central bank at the risk-free interest rate, R^f_{t+1}, then funding costs as a share of GDP would increase by

$$\text{fcr}_t \equiv \frac{R^f_{t+1} - R^l_{t+1}}{R^f_{t+1}} \frac{n_{t+1}(1 - \zeta_{t+1})}{\text{GDP}_t}.$$

Next, we calculate fcr_t based on quarterly U.S. data for the period since 1999.

The top panel of figure 5 displays the (inflation adjusted) gross reserves rate, R^r_{t+1}, the gross risk-free rate, R^f_{t+1}, and the gross deposit rate, R^n_{t+1}; the data spans the interval 1999q1–2021q1. To construct these series we use FRED data and Kurlat's (2019) estimates of the risk-free, "illiquid" interest rate as well as the deposit rate. Appendix G contains detailed information.

[73] Piazzesi and Schneider (2021) model CBDC as loanable funds and they assume that drawing a credit line creates deposits by credit.

[74] Related, Benigno et al. (2022) analyze how the international adoption of a CBDC interacts with national asset pricing conditions. See Ferrari Minesso et al. (2022) for a quantitative analysis in a DSGE model.

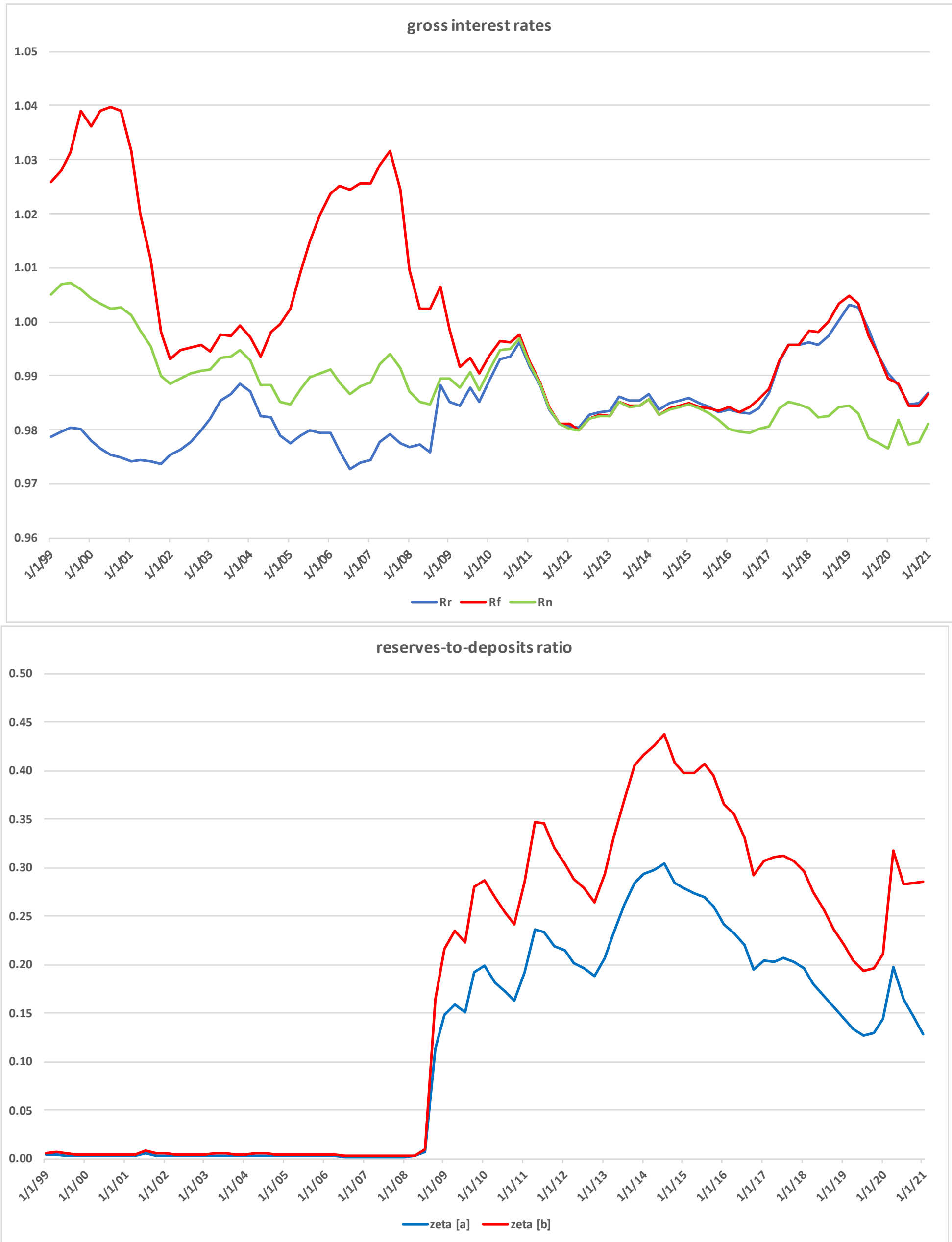


Figure 5: Gross interest rates (top) and two measures of the reserves-to-deposits ratio (bottom).

The figure shows that the (inflation adjusted) interest rate on reserves fluctuates in a band between -3 and 0.5 percent while the risk-free rate varies between -2 and 4 percent. After 2010, the two rates nearly coincide. The deposit rate lies between the reserves rate and the risk-free rate before the financial crisis and below the two other rates at the end of the sample. In the first half of the 2010s, there are only tiny spreads between the three rates.

The bottom panel of figure 5 displays the reserves-to-deposits ratio, ζ_{t+1}, over the same period. We use FRED data as well as data constructed by Lucas and Nicolini (2015) for reserves and deposits. For the deposit series, we use two alternative measures. The first, indicated by [a], represents the sum of checkable and savings deposits. The second, indicated by [b], represents the sum of checkable deposits and money market deposit accounts as specified by Lucas and Nicolini (2015). Appendix G contains more information.

Irrespective of the exact measure, the reserves-to-deposits ratio strongly increases in mid 2008, from a very low level (at which it had been since the early 1980s). It reaches a maximum of 30 or 45 percent, depending on the measure, in mid 2014 and falls afterwards before increasing again at the beginning of the COVID-19 crisis.

Figures 6 and 7 display the implied R^l_{t+1} and fcr_t. In each case we report the results for either measure of the reserves-to-deposits ratio; the differences are minor. In figure 6 we compute R^l_{t+1} under the assumption that $\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t = 0.01$, in line with estimates of banks' operating costs.[75] In figure 7 we infer $\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t$ from the data under the assumption that banks behave according to the model.

Consider first figure 6. The equivalent loan rate falls from nearly 2 percent early in the sample to -1 percent at the end, with a temporary increase to 1 percent in late 2010. We can distinguish several phases:

- Prior to 2008 R^l_{t+1} roughly equals $R^n_{t+1} + 0.01$ as the reserves-to-deposits ratio is tiny and $R^f_{t+1} \approx 1$ (see equation (18)).

- In 2008 the reserves-to-deposits ratio strongly increases; one dollar of funding for capital investment now requires substantially more than one dollar of deposits. Since the deposit rate exceeds the interest rate on reserves, the equivalent loan rate rises.

- Between 2009 and 2015, R^l_{t+1} follows the interest rate on reserves—due to the strong compression of interest rates—plus a term that reflects operating and liquidity substitution costs and the reserves-to-deposits ratio.

- Finally, after 2015 the reduction of the deposit rate relative to the reserves rate contributes negatively to R^l_{t+1}.

The bottom panel of figure 6 illustrates the implied funding-cost-reduction-at-risk. The time series fluctuates between -0.7 and 0.8 percent of GDP, reflecting several drivers: The long-term decline in $R^f_{t+1} - R^l_{t+1}$ and fluctuations around this trend; the U-shaped path of $1-\zeta_{t+1}$ after 2008; and an increasing deposits-to-GDP ratio, in particular towards the end of the sample. Again, we can distinguish several phases:

- In the beginning of the sample, money creation by banks reduces their funding costs by roughly 0.5 percent of GDP because the risk-free rate exceeds the costs of deposit funding. Equivalently, banks benefit from the equivalent central bank loan in the money-based system because the risk-free rate exceeds the equivalent loan rate.

[75] Lucas and Nicolini (2015, p. 57) assume that banks' costs of check processing equal 1 percent of GDP. Philippon (2015) estimates that the costs of financial intermediation equal 1.5 to 2 percent of intermediated assets.

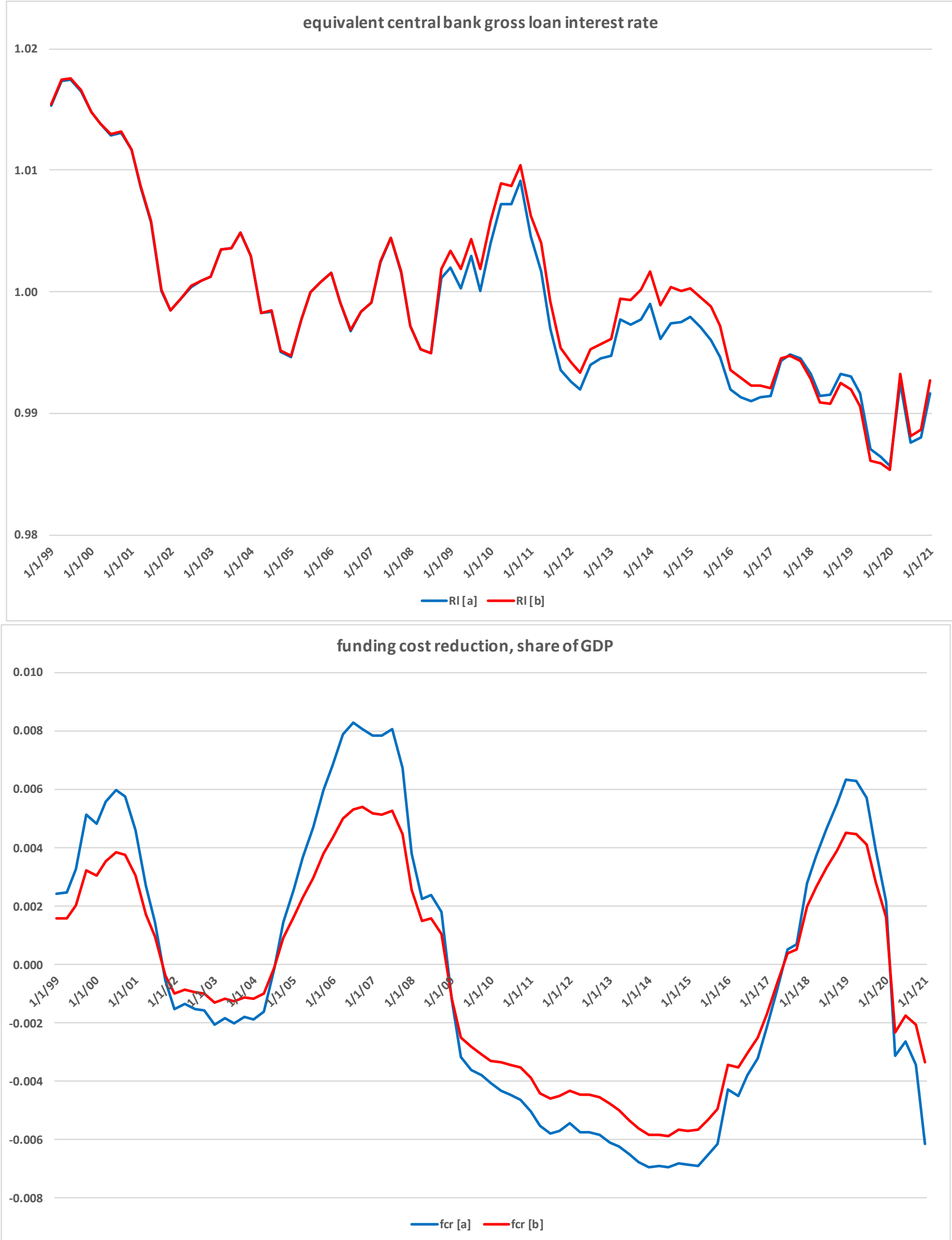


Figure 6: Two measures of the equivalent central bank gross loan interest rate (top) and banks' funding-cost-reduction-at-risk (bottom) under the assumption that $\nu + \omega_t - \theta_t = 0.01$.

- From 2002 to 2004 the risk-free rate is low and so is the spread between the risk-free rate and the equivalent loan rate. As a consequence fcr_t is negative.
- In 2005 and 2006 the risk-free rate rises again, pushing fcr_t to $0.5 - 0.8$ percent. In the following two years this effects weakens.
- From 2008 the three market interest rates converge and between 2011 and 2015 they

practically coincide; fcr_t therefore increasingly mimics $-(\nu+\omega_t-\theta_t)n_{t+1}/\text{GDP}_t$ (see equation (18) and the formula for fcr_t).

- From 2016 to 2019 the fall in the equivalent loan rate and the rise in the risk-free rate contribute to an increase in fcr_t.
- At the end of the sample the fall in the risk-free rate relative to the equivalent loan rate reduces fcr_t and pushes it back into negative territory.

Turn next to figure 7. Rather than positing $\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1}) - \theta_t = 0.01$ we now infer operating and liquidity substitution costs from the data assuming that banks set the deposit rate according to the model. Specifically, imposing assumption 1 and using the bank's first-order condition $\nu+\omega_t(\zeta_{t+1}, \zeta_{t+1})-\theta_t = (1-\psi)\chi^n_{t+1}-\zeta_{t+1}\chi^r_{t+1}$ (see appendix B), we find the simple expressions

$$\begin{aligned} R^l_{t+1} &= R^f_{t+1} - \frac{\psi}{1-\zeta_{t+1}}(R^f_{t+1} - R^n_{t+1}), \\ \text{fcr}_t &= \psi\chi^n_{t+1}\frac{n_{t+1}}{\text{GDP}_t}, \end{aligned}$$

according to which stronger market power (higher ψ) reduces the equivalent loan rate and increases the funding-cost-reduction-at-risk. For our calculations we let $\psi = 0.5$.

The equivalent loan rate again displays a downward trend, falling by roughly 2 percent over the sample period with a temporary reversal around the time of the financial crisis. The funding-cost-reduction-at-risk varies between 0 and 2 percent of GDP. It exhibits the same cyclical behavior as in figure 6 but is always positive. This is a consequence of the fact that R^f_{t+1} always exceeds R^n_{t+1} such that $\chi^n_{t+1} > 0$.

G Data

We use the quarterly average of the FRED series `IOER` (2008q4–2021q1) for the nominal interest rate on reserves. To compute the gross real interest rate we divide the gross nominal rate by the gross inflation rate, Π_{t+1} (see below). Since no interest on reserves was paid prior to 2008[76] we set $R^r_{t+1} = 1/\Pi_{t+1}$ prior to 2008.

We use the quarterly average of the FRED series `CPILFESL_PC1` (Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average, Percent Change from Year Ago, Index 1982–1984=100, Seasonally Adjusted) for gross inflation.

We use quarterly averages of Kurlat's (2019) monthly estimates (1999m01–2017m12) of the risk-free rate and the deposit rate.[77] Kurlat (2019) provides two estimates of the latter (based on RateWatch data), one based on data for checking accounts and the other for money market accounts. We compute a weighted average of the two estimates where the weights correspond to the relative size of checkable and savings deposits (see below). We adjust the constructed interest rate series using the inflation series defined before. Since

[76] See `https://www.federalreserve.gov/monetarypolicy/reqresbalances.htm`.

[77] We thank Pablo Kurlat for sharing his data.

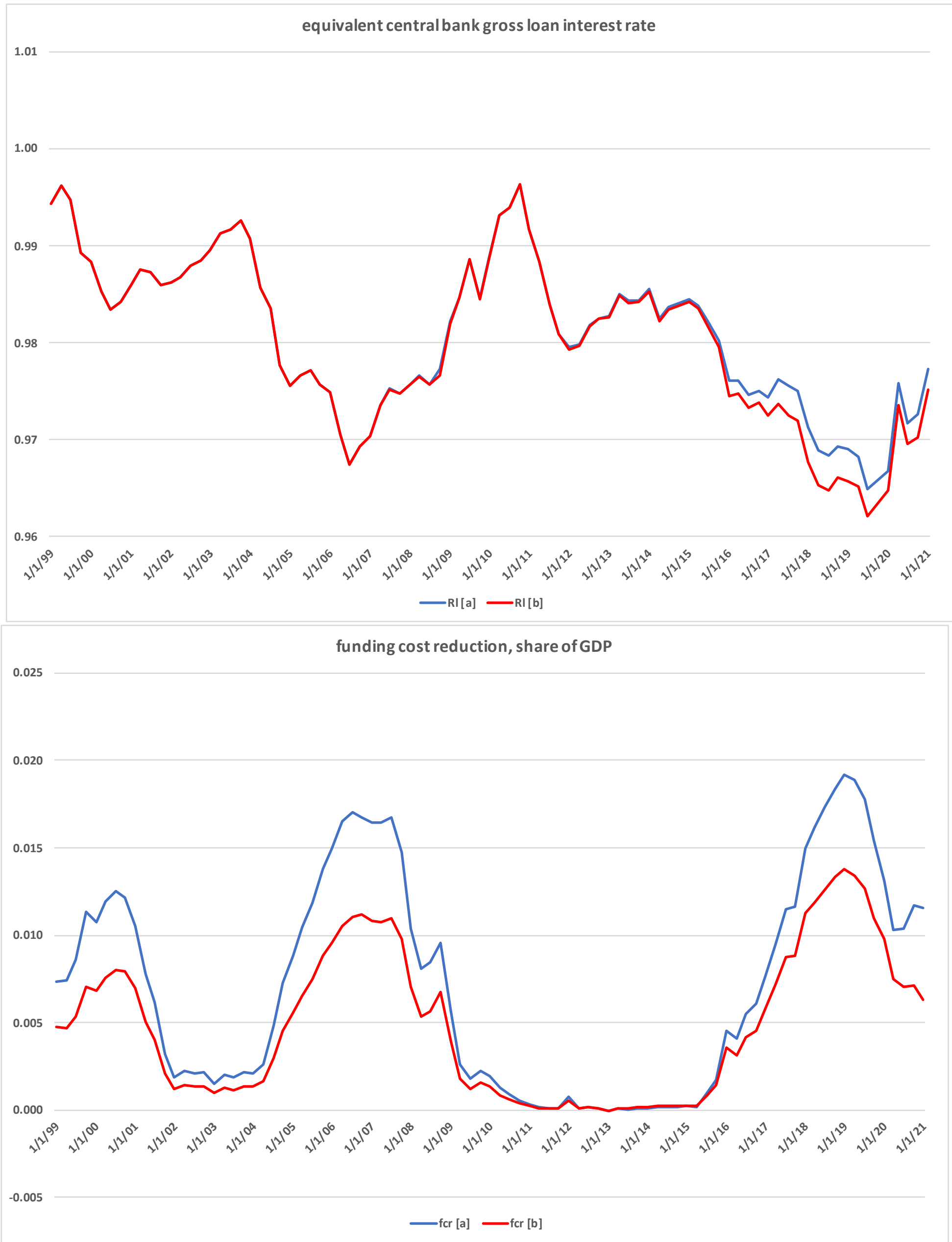


Figure 7: Two measures of the equivalent central bank gross loan interest rate (top) and banks' funding-cost-reduction-at-risk (bottom) under assumption 1 with $\psi = 0.5$.

Kurlat's (2019) series end in 2017 we extrapolate them at the end of the sample using projections on the quarterly averages of the FRED series TB3MS (3-Month Treasury Bill: Secondary Market Rate, Percent, Not Seasonally Adjusted) and AAA (Moody's Seasoned Aaa Corporate Bond Yield, Percent, Not Seasonally Adjusted).

We use quarterly averages of the FRED series TCDSL (Total Checkable Deposits, Billions of Dollars, Seasonally Adjusted) and SAVINGSL (Savings Deposits - Total, Billions of Dollars, Seasonally Adjusted) for checkable and savings deposits, respectively. Since the series TCDSL and SAVINGSL were discontinued after 2020q4 and 2020q1, respectively, we extrapolate them at the end of the sample using projections on the quarterly average of the FRED series DEMDEPSL (Demand Deposits: Total, Billions of Dollars, Seasonally Adjusted).

We use the quarterly average of the FRED series RESBALNS (Total Reserve Balances Maintained with Federal Reserve Banks, Billions of Dollars, Not Seasonally Adjusted) for reserves. Since the series RESBALNS was discontinued after 2020q2 we use the quarterly average of the FRED series BOGMBBM (Reserve Balances, Millions of Dollars, Not Seasonally Adjusted; divided by thousand) for the most recent periods.

We use two alternative series for deposits. The first series ([a]) is the sum of the quarterly averages of the FRED series TCDSL and SAVINGSL defined before. The second series ([b]) is the sum of the quarterly average of the FRED series TCDSL and the quarterly money market deposit account (MMDA) series constructed by Lucas and Nicolini (2015).[78] Since the updated MMDA series ends in 2020q2 we extrapolate it at the end of the sample using projections on the cumulative sums of the quarterly FRED series HNOMMFQ027S (Households and Nonprofit Organizations; Money Market Fund Shares; Asset, Flow, Millions of Dollars, Seasonally Adjusted; divided by thousand), BOGZ1FA103034000Q (Nonfinancial Corporate Business; Money Market Fund Shares; Asset, Flow, Millions of Dollars, Seasonally Adjusted; divided by thousand), and NNBMMFQ027S (Nonfinancial Noncorporate Business; Money Market Fund Shares; Asset, Flow, Millions of Dollars, Seasonally Adjusted; divided by thousand) as well as lags of the series.

We compute ζ_{t+1} as the ratio of the reserve series and either of the two deposit series. We compute deposits as a share of GDP as the ratio of either of the two deposit series and the FRED series GDP (Gross Domestic Product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate).

We use the quarterly FRED series A587RC1Q027SBEA (Corporate profits with inventory valuation and capital consumption adjustments: Domestic industries: Financial) for financial sector profits.

H Optimality Under Assumptions 1 and 2

H.1 Liquidity Provision by the Central Bank

When the social planner opts for money the liquidity premium under the Ramsey policy equals

$$\chi_{t+1}^{m\star} = \mu,$$

see condition (RA-1) in the text.

[78] We thank Luca Benati for sharing the MMDA series.

H.2 Liquidity Provision by Banks

H.2.1 Liquidity Substitution Costs and Spreads Under Assumption 2

When the social planner opts for deposits then the optimal reserves-to-deposits ratio, $\zeta_{t+1}^{\star}$, solves the first-order condition $\omega_{1,t}(\zeta_{t+1}, \zeta_{t+1}) + \omega_{2,t}(\zeta_{t+1}, \zeta_{t+1}) = -\rho$, i.e.,

$$\phi_t(\varphi + \bar{\varphi})(1 - \zeta_{t+1})^{\varphi+\bar{\varphi}-1} = \rho$$

such that

$$\zeta_{t+1}^{\star} = 1 - \left(\frac{\rho}{\phi_t(\varphi + \bar{\varphi})}\right)^{\frac{1}{\varphi+\bar{\varphi}-1}}, \quad \omega_t(\zeta_{t+1}^{\star}, \zeta_{t+1}^{\star}) = \phi_t \left(\frac{\rho}{\phi_t(\varphi + \bar{\varphi})}\right)^{\frac{\varphi+\bar{\varphi}}{\varphi+\bar{\varphi}-1}}.$$

(We assume that ρ, ϕ_t, φ, and $\bar{\varphi}$ are such that $\zeta_{t+1}^{\star} \in (0, 1)$.) The optimal shadow liquidity premium on reserves thus equals

$$-\omega_{1,t}(\zeta_{t+1}^{\star}, \zeta_{t+1}^{\star}) = \rho \frac{\varphi}{\varphi + \bar{\varphi}},$$

which is smaller than ρ if and only if reserves generate externalities ($\bar{\varphi} > 0$).

The optimal spread on deposits under the Ramsey policy is given by

$$\chi_{t+1}^{n\star} = \nu + \omega_t(\zeta_{t+1}^{\star}, \zeta_{t+1}^{\star}) + \zeta_{t+1}^{\star}\rho,$$

see condition (RA-2) in the main text.

H.2.2 Deposit Subsidy Under Assumption 1

When the central bank targets ξ_{t+1} or $m_{t+1} = 0$ the equilibrium spread on deposits satisfies equation (8a'). Equilibrium bank choices thus are socially optimal if

$$\chi_{t+1}^{n\star} = \nu + \omega_t(\zeta_{t+1}^{\star}, \zeta_{t+1}^{\star}) + \zeta_{t+1}^{\star}\rho = \frac{\nu + \tilde{\omega}_t(-\chi_{t+1}^{r\star}) - \theta_t^{\star}}{1 - \psi}$$

or, using the fact that $\tilde{\omega}_t(-\chi_{t+1}^{r\star}) = \omega_t(\zeta_{t+1}^{\star}, \zeta_{t+1}^{\star}) + \chi_{t+1}^{r\star}\zeta_{t+1}^{\star}$,

$$\theta_t^{\star} = \psi \left(\nu + \omega_t(\zeta_{t+1}^{\star}, \zeta_{t+1}^{\star}) + \zeta_{t+1}^{\star}\rho\right) - \zeta_{t+1}^{\star}(\rho - \chi_{t+1}^{r\star}).$$

References

Andolfatto, D. (2021). Assessing the impact of central bank digital currency on private banks, *Economic Journal* **131**(634): 525–540.

Auclert, A., Bardóczy, B., Rognlie, M. and Straub, L. (2021). The sequence-space Jacobian to solve and estimate heterogeneous-agent models, *Econometrica* **89**(5): 2375–2408.

Auer, R. and Böhme, R. (2020). The technology of retail central bank digital currency, *BIS Quarterly Review* pp. 85–100.

Barrdear, J. and Kumhof, M. (2022). The macroeconomics of central bank digital currencies, *Journal of Economic Dynamics and Control* **forthcoming**.

Barro, R. J. (1974). Are government bonds net wealth?, *Journal of Political Economy* **82**(6): 1095–1117.

Begenau, J. (2020). Capital requirements, risk choice, and liquidity provision in a business cycle model, *Journal of Financial Economics* **136**(32): 355–378.

Benes, J. and Kumhof, M. (2012). The Chicago plan revisited, *Working Paper 12/202*, International Monetary Fund, Washington.

Benigno, P., Schilling, L. and Uhlig, H. (2022). Cryptocurrencies, currency competition, and the impossible trinity, *Journal of International Economics* **136**: 103601.

Bernanke, B. S. and Blinder, A. S. (1988). Credit, money, and aggregate demand, *American Economic Review* **78**(2): 435–439.

Bianchi, J. and Bigio, S. (2020). Banks, liquidity management, and monetary policy. Unpublished, Federal Reserve Bank of Minneapolis and UCLA.

Bindseil, U. (2020). Tiered CBDC and the financial system, *Working Paper 2351*, European Central Bank, Frankfurt.

BIS (1996). Implications for central banks of the development of electronic money, *Report*, BIS, Basel.

BIS (2018). Central bank digital currencies, *Report*, BIS: Committee on Payments and Market Infrastructures, and Markets Committee, Basel.

Boar, C. and Wehrli, A. (2021). Ready, steady, go? – Results of the third BIS survey on central bank digital currency, *BIS Papers 114*, BIS, Basel.

Board of Governors (2022). Money and payments: The U.S. dollar in the age of digital transformation, *Research and Analysis*, Federal Reserve Board, Washington, DC.

Bolton, P., Li, Y., Wang, N. and Yang, J. (2020). Dynamic banking and the value of deposits, *Working Paper 2020-13*, Ohio State University, Fisher College of Business, Columbus.

Bordo, M. D. and Levin, A. T. (2017). Central bank digital currency and the future of monetary policy, *Working Paper 23711*, NBER, Cambridge, Massachusetts.

Böser, F. and Gersbach, H. (2020). Monetary policy with a central bank digital currency: The short and the long term, *Discussion Paper 15322*, CEPR, London.

Brunnermeier, M. K., Merkel, S. A. and Sannikov, Y. (2022). Debt as safe asset, *Working Paper 29626*, NBER, Cambridge, Massachusetts.

Brunnermeier, M. K. and Niepelt, D. (2019). On the equivalence of private and public money, *Journal of Monetary Economics* **106**: 27–41.

Bryant, J. (1983). Government irrelevance results: A simple exposition, *American Economic Review* **73**(4): 758–761.

Buiter, W. H. (2009). Negative nominal interest rates: Three ways to overcome the zero lower bound, *Working Paper 15118*, NBER, Cambridge, Massachusetts.

Bullard, J. and Mitra, K. (2002). Learning about monetary policy rules, *Journal of Monetary Economics* **49**(6): 1105–1129.

Burlon, L., Montes-Galdón, C., Muñoz, M. A. and Smets, F. (2022). The optimal quantity of CBDC in a bank-based economy, *Discussion Paper 16995*, CEPR, London.

Chamley, C. and Polemarchakis, H. (1984). Assets, general equilibrium and the neutrality of money, *Review of Economic Studies* **51**(1): 129–138.

Chari, V. V. and Phelan, C. (2014). On the social usefulness of fractional reserve banking, *Journal of Monetary Economics* **65**(C): 1–13.

Chiu, J., Davoodalhosseini, M., Jiang, J. and Zhu, Y. (2019). Bank market power and central bank digital currency: Theory and quantitative assessment, *Staff Working Paper 2019-20*, Bank of Canada, Ottawa.

Croushore, D. (1993). Money in the utility function: Functional equivalence to a shoppingtime model, *Journal of Macroeconomics* **15**(1): 175–182.

Cúrdia, V. and Woodford, M. (2009). Credit frictions and optimal monetary policy. Unpublished, Federal Reserve Bank of New York and Columbia University.

Di Tella, S. (2020). Risk premia and the real effects of money, *American Economic Review* **110**(7): 1995–2040.

Diamond, D. W. and Rajan, R. G. (2011). Fear of fire sales, illiquidity seeking, and credit freezes, *Quarterly Journal of Economics* **126**(2): 557–591.

Divino, J. A. and Orrillo, J. (2017). Failure of the Ricardian equivalence theory in economies with incomplete markets, *Brazilian Review of Econometrics* **37**(1): 19–30.

Drechsler, I., Savov, A. and Schnabl, P. (2017). The deposits channel of monetary policy, *Quarterly Journal of Economics* **132**(4): 1819–1876.

Drechsler, I., Savov, A. and Schnabl, P. (2021). Banking on deposits: Maturity transformation without interest rate risk, *Journal of Finance* **76**(3).

Faure, S. and Gersbach, H. (2018). Money creation in different architectures, *Discussion Paper 13156*, CEPR.

Feenstra, R. C. (1986). Functional equivalence between liquidity costs and the utility of money, *Journal of Monetary Economics* **17**(2): 271–291.

Ferrari Minesso, M., Mehl, A. and Stracca, L. (2022). Central bank digital currency in an open economy, *Journal of Monetary Economics* **127**: 54–68.

Fisher, I. (1935). *100% Money*, Adelphi, New York.

Freixas, X. and Rochet, J.-C. (2008). *Microeconomics of Banking*, 2. edn, MIT Press, Cambridge, Massachusetts.

Friedman, M. (1969). The optimum quantity of money, *in* M. Friedman (ed.), *The Optimum Quantity of Money and Other Essays*, Aldine, Chicago, chapter 1, pp. 1–50.

Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle*, 2. edn, Princeton University Press, Princeton.

Garratt, R. J. and Lee, M. J. (2021). Monetizing privacy, *Staff Report 958*, Federal Reserve Bank of New York, New York.

Garratt, R. J. and van Oordt, M. R. C. (2021). Privacy as a public good: A case for electronic cash, *Journal of Political Economy* **129**(7): 2157–2180.

Garratt, R. J. and Zhu, H. (2021). On interest-bearing central bank digital currency with heterogeneous banks. Unpublished, University of California Santa Barbara.

Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy, *Journal of Monetary Economics* **58**(1): 17–34.

Goodfriend, M. and King, R. G. (1997). The new neoclassical synthesis and the role of monetary policy, *in* B. S. Bernanke and J. J. Rotemberg (eds), *NBER Macroeconomics Annual 1997*, Vol. 12, MIT Press, Cambridge, Massachusetts, pp. 231–283.

Gurley, J. G. and Shaw, E. S. (1960). *Money in a Theory of Finance*, Brookings Institution, Washington.

Hanson, S. G., Shleifer, A., Stein, J. C. and Vishny, R. W. (2015). Banks as patient fixed-income investors, *Journal of Financial Economics* **117**(3): 449–469.

He, Z. and Krishnamurthy, A. (2018). Intermediary asset pricing and the financial crisis, *Annual Review of Financial Economics* **10**(1): 173–197.

Jackson, T. and Pennacchi, G. (2021). How should governments create liquidity?, *Journal of Monetary Economics* **118**: 281–295.

Kahn, C. M., McAndrews, J. and Roberds, W. (2005). Money is privacy, *International Economic Review* **46**(2): 377–399.

Kahn, C. M., Rivadeneyra, F. and Wong, T.-N. (2018). Should the central bank issue e-money?, *Staff Working Paper 2018-58*, Bank of Canada, Ottawa.

Kashyap, A. K., Rajan, R. and Stein, J. C. (2002). Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking, *Journal of Finance* **57**(1): 33–73.

Keister, T. and Monnet, C. (2020). Central bank digital currency: Stability and information. Unpublished, Rutgers University and University of Bern.

Keister, T. and Sanches, D. (2022). Should central banks issue digital currency?, *Review of Economic Studies* **forthcoming**.

King, M. (2016). *The End of Alchemy: Money, Banking, and the Future of the Global Economy*, Norton, New York.

King, R. G. and Wolman, A. L. (1996). Inflation targeting in a St. Louis model of the 21st century, *Federal Reserve Bank of St. Louis Review* **78**(3): 83–107.

Kiyotaki, N. and Moore, J. (2019). Liquidity, business cycles, and monetary policy, *Journal of Political Economy* **127**(6): 2926–2966.

Klein, M. A. (1971). A theory of the banking firm, *Journal of Money, Credit, and Banking* **3**(2): 205–218.

Knight, F. H., Cox, G. V., Director, A., Douglas, P. H., Fisher, I., Hart, A. G., Mints, L. W., Schultz, H. and Simons, H. C. (1933). Memorandum on banking reform. Henry C. Wallace papers.

Krishnamurthy, A. and Vissing-Jorgensen, A. (2012). The aggregate demand for treasury debt, *Journal of Political Economy* **120**(2): 233–267.

Kumhof, M. and Noone, C. (2021). Central bank digital currencies—design principles for financial stability, *Economic Analysis and Policy* **71**: 553–572.

Kurlat, P. (2019). Deposit spreads and the welfare cost of inflation, *Journal of Monetary Economics* **106**: 78–93.

Lagos, R. and Zhang, S. (2020). The limits of onetary economics: On money as a latent medium of exchange, *Working Paper 26756*, NBER, Cambridge, Massachusetts.

Lucas, R. E. and Nicolini, J. P. (2015). On the stability of money demand, *Journal of Monetary Economics* **73**: 48–65.

McCallum, B. T. and Goodfriend, M. S. (1987). Demand for money: Theoretical studies, *in* J. Eatwell, P. Newman and M. Milgate (eds), *The New Palgrave: A Dictionary of Economics*, Macmillan Press, London, pp. 775–781.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment, *American Economic Review* **48**(3): 261–297.

Monti, M. (1972). Deposit, credit and interest rate determination under alternative bank objective functions, *in* K. Shell and G. P. Szegö (eds), *Mathematical Methods in Investment and Finance*, North-Holland, Amsterdam, pp. 431–454.

Nagel, S. (2016). The liquidity premium of near-money assets, *Quarterly Journal of Economics* **131**(4): 1927–1971.

Niepelt, D. (2020a). Monetary policy with reserves and CBDC: Optimality, equivalence, and politics, *Discussion Paper 15457*, CEPR.

Niepelt, D. (2020b). Reserves for all? Central Bank Digital Currency, deposits, and their (non)-equivalence, *International Journal of Central Banking* **16**(3): 211–238.

Niepelt, D. (2021). Reserves for all: Political rather than macroeconomic risks, *in* D. Niepelt (ed.), *Central Bank Digital Currency: Considerations, Projects, Outlook*, CEPR Press, London, chapter 5, pp. 39–44.

Ozdenoren, E., Yuan, K. and Zhang, S. (2021). Safe assets as balance sheet multipliers. Unpublished, LBS and LSE.

Parlour, C. A., Rajan, U. and Walden, J. (2022). Payment system externalities, *Journal of Finance* **77**(2): 1019–1053.

Philippon, T. (2015). Has the US finance industry become less efficient? On the theory and measurement of financial intermediation, *American Economic Review* **105**(4): 1408–1438.

Phillips, C. A. (1920). *Bank Credit*, McMillan, New York.

Piazzesi, M. and Schneider, M. (2021). Credit lines, bank deposits or CBDC? competition and efficiency in modern payment systems. Unpublished, Stanford University.

Pulley, L. B. and Humphrey, D. B. (1993). The role of fixed costs and cost complementarities in determining scope economies and the cost of narrow banking proposals, *Journal of Business* **66**(3): 437–462.

Rocheteau, G. and Nosal, E. (2017). *Money, Payments, and Liquidity*, 2. edn, MIT Press, Cambridge, Massachusetts.

Sargent, T. J. (1987). *Dynamic Macroeconomic Theory*, Harvard University Press, Cambridge, Massachusetts.

Saving, T. R. (1971). Transactions costs and the demand for money, *American Economic Review* **61**(3): 407–420.

Schilling, L., Fernández-Villaverde, J. and Uhlig, H. (2020). Central bank digital currency: When price and bank stability collide, *Working Paper 28237*, NBER, Cambridge, Massachusetts.

Schmiedel, H., Kostova, G. and Ruttenberg, W. (2012). The social and private costs of retail payment instruments: A European perspective, *Occasional Paper 137*, European Central Bank, Frankfurt.

Shleifer, A. and Vishny, R. W. (1992). Liquidation values and debt capacity: A market equilibrium approach, *Journal of Finance* **47**(4): 1343–1366.

Shleifer, A. and Vishny, R. W. (2011). Fire sales in finance and macroeconomics, *Journal of Economic Perspectives* **25**(1): 29–48.

Sidrauski, M. (1967). Rational choice and patterns of growth in a monetary economy, *American Economic Review* **57**(2): 534–544.

Stein, J. C. (2012). Monetary policy as financial stability regulation, *Quarterly Journal of Economics* **127**(1): 57–97.

Taudien, R. (2020). Borrowing costs, liquidity and full reserve banking. Thesis paper, University of Bern.

Taylor, J. B. (1999). A historical analysis of monetary policy rules, *in* J. B. Taylor (ed.), *Monetary Policy Rules*, University of Chicago Press, Chicago, chapter 7, pp. 319–348.

Tobin, J. (1963). Commercial banks as creators of "money", *Discussion Paper 159*, Cowles Foundation, New Haven.

Tobin, J. (1969). A general equilibrium approach to monetary theory, *Journal of Money, Credit, and Banking* **1**(1): 15–29.

Tobin, J. (1985). Financial innovation and deregulation in perspective, *Bank of Japan Monetary and Economic Studies* **3**(2): 19–29.

Ulate, M. (2021). Going negative at the zero lower bound: The effects of negative nominal interest rates, *American Economic Review* **111**(1): 1–40.

Van den Heuvel, S. J. (2008). The welfare cost of bank capital requirements, *Journal of Monetary Economics* **55**(2): 298–320.

Van den Heuvel, S. J. (2019). The welfare effects of bank liquidity and capital requirements. Unpublished, Federal Reserve Board.

Vandeweyer, Q. (2019). Treasury debt and the pricing of short-term assets. Unpublished, European Central Bank and Sciences Po.

Wallace, N. (1981). A Modigliani-Miller theorem for open-market operations, *American Economic Review* **71**(3): 267–274.

Walsh, C. E. (2017). *Monetary Theory and Policy*, 4. edn, MIT Press, Cambridge, Massachusetts.

Wang, P. and Yip, C. K. (1992). Alternative approaches to money and growth, *Journal of Money, Credit, and Banking* **24**(4): 553–562.

Wang, Y., Whited, T. M., Wu, Y. and Xiao, K. (2020). Bank market power and monetary policy transmission: Evidence from a structural estimation, *Working Paper 27258*, NBER, Cambridge, Massachusetts.

Williamson, S. (2019). Central bank digital currency: Welfare and policy implications. Unpublished, University of Western Ontario.

Woodford, M. (2001). Fiscal requirements for price stability, *Journal of Money, Credit, and Banking* **33**(3): 669–728.

Woodford, M. (2003). *Interest and Prices*, Princeton University Press, Princeton.

Yun, T. (1996). Nominal price rigidity, money supply endogeneity, and business cycles, *Journal of Monetary Economics* **37**(2): 345–370.