DISCUSSION PAPER SERIES

DP18444

MONEY AND BANKING WITH RESERVES AND CBDC

Dirk Niepelt

MONETARY ECONOMICS AND FLUCTUATIONS AND BANKING AND CORPORATE FINANCE

CEPR

ISSN 0265-8003

MONEY AND BANKING WITH RESERVES AND CBDC

Dirk Niepelt

Discussion Paper DP18444
Published 09 September 2023
Submitted 06 September 2023

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

MONEY AND BANKING WITH RESERVES AND CBDC

Abstract

We analyze the role of retail central bank digital currency (CBDC) and reserves when banks exert deposit market power and liquidity transformation entails externalities. Optimal monetary architecture minimizes the social costs of liquidity provision and optimal monetary policy follows modified Friedman (1969) rules. Interest rates on reserves and CBDC should differ. Calibrations robustly suggest that CBDC provides liquidity more efficiently than deposits unless the central bank must refinance banks and this is very costly. Accordingly, the optimal share of CBDC in payments tends to exceed that of deposits.

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

Keywords: Central bank digital currency

Dirk Niepelt - dirk.niepelt@vwi.unibe.ch
University Of Bern and CEPR

Money and Banking with Reserves and CBDC*

Dirk Niepelt†

September 6, 2023

Abstract

We analyze the role of retail central bank digital currency (CBDC) and reserves when banks exert deposit market power and liquidity transformation entails externalities. Optimal monetary architecture minimizes the social costs of liquidity provision and optimal monetary policy follows modified Friedman (1969) rules. Interest rates on reserves and CBDC should differ. Calibrations robustly suggest that CBDC provides liquidity more efficiently than deposits unless the central bank must refinance banks and this is very costly. Accordingly, the optimal share of CBDC in payments tends to exceed that of deposits.

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

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

*For valuable comments and suggestions, we are grateful to Urban Jermann, an associate editor and anonymous referees; 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). Pablo Kurlat shared and Luca Benati, Han Gao and Juan Pablo Nicolini helped with data used in a previous version of the paper.

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

1 Introduction

The prospect of retail central bank digital currency (CBDC) rekindles interest in the architecture of the monetary system. In modern economies, this architecture exhibits two tiers: Nonbanks transact using bank deposits or claims on deposits, and banks settle payments with central-bank issued reserves. New financial service providers build on the existing payment rails but do not undermine this basic architecture. CBDC, in contrast, subverts the two-tier system by offering the general public a digital, central-bank issued payment instrument that directly competes with deposits.[1]

Should CBDC be introduced and how should it be remunerated? To answer these questions we analyze how monetary architecture shapes the social costs of liquidity provision. We compare a two-tier payment system like the current one, a single-tier CBDC based system, and a hybrid architecture in which reserves, deposits and CBDC co-exist.

We find that a two-tier system generates lower direct costs of payments than a CBDC based architecture that resembles narrow banking (reserves equal deposits). Another downside of CBDC results from its crowding out of deposits: While the central bank can offset deposit outflows by lending to banks, this is costly due to agency frictions. The upside of CBDC is that it allows to bypass frictions in the banking sector such as deposit market power and externalities from reserve holdings. The social costs of addressing these frictions, which CBDC avoids, erode the advantage of the two-tier system.

Calibrations robustly suggest that CBDC provides liquidity more efficiently than deposits unless deposit outflows from banks must be offset by central bank loans that generate large social costs. Optimal portfolios reflect this CBDC cost advantage: The welfare-maximizing share of CBDC in payments is generally larger than that of deposits. Lower substitutability between CBDC and deposits implies more balanced shares.

The remuneration of CBDC and reserves should be set to allocate liquidity efficiently. This requires different interest rates on the two central bank liabilities, reflecting their distinct operating costs and externalities. Interest on CBDC may also be set to check banks' price setting power.

We derive these results in the textbook growth and business cycle model, which we augment with liquidity demand, banks and multiple costly payment instruments. Households value the liquidity of deposits and CBDC. Banks exert deposit market power and invest in capital; they also hold reserves to avoid costly liquidity shortages but do not

[1] See for instance Board of Governors (2022) and Boar and Wehrli (2021). Noncash public retail payment instruments existed in the past when central banks offered accounts for nonbanks (BIS, 2018). CBDC 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).

fully internalize the benefits. The central bank issues CBDC and reserves and controls their price or quantity; it may also subsidize deposits to counteract monopsony distortions. Cash and government bonds play unimportant roles. Adding lending frictions to this liquidity centric setup does not change the results as we explain.

In the conventional two-tier system, the central bank's choice of deposit subsidies and interest on reserves affects the cost structure of banks, their price setting and thus the liquidity premium on deposits. In a CBDC based system, the transmission mechanism is more direct since the central bank sets the spread that households face. 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.

Our normative analysis starts with a characterization of the social planner outcome. The planner provides payment instruments up to the point where the marginal liquidity benefit equals the marginal social cost; that is, it follows a generalized version of the Friedman (1969) rule that accounts for the resource costs of payment operations. As for the source of liquidity, the planner relies on the monetary architecture that generates the lowest effective resource costs. Depending on household preferences and payment technology this may be a fully CBDC based, deposit and reserves based, or hybrid system.

The Ramsey government implements the planner outcome. In a CBDC based system, the central bank sets the liquidity premium on CBDC at the level that reflects CBDC operating costs. In a two-tier system, the situation is more challenging since the government needs to correct two distortions in the banking sector, but this is feasible: The optimal interest rate on reserves includes a Pigou (1920) subsidy such that banks internalize the reserve externality; and the optimal deposit subsidy induces banks to charge the efficient liquidity premium on deposits. In a hybrid system, the central bank employs all these instruments.

More bank market power requires a higher deposit subsidy to encourage banks to lengthen their balance sheets. But importantly, stronger reserve externalities call for a lower deposit subsidy. Intuitively, when the government raises interest on reserves because of their external benefits, it sterilizes the effect on the deposit margin by lowering the deposit subsidy.

Under the optimal policy, deposits are only partly backed by reserves because starting from full backing, liquidity transformation saves banks costs at the margin. Accordingly, CBDC based payments must be more efficient than a narrow-bank arrangement to be competitive with deposit based payments. The calibrated model suggests that this is not a high hurdle to overcome but sufficient to make the case for CBDC a weak one if we strictly focus on payment operations.

The ranking becomes more clear-cut when we allow for an important bank lending channel, that is, more productive capital investment by banks than the central bank. Since the central bank can pass CBDC funds on to banks by lending to them and since this decouples liquidity provision from bank lending (Brunnermeier and Niepelt, 2019) the channel does not undermine our liquidity centric perspective. But it further weakens the case for CBDC if pass-through funding carries social costs, as we argue it does.

CBDC gains in appeal, however, when we also account more fully for the social costs of the two-tier architecture. We consider how information frictions, lack of commitment and too-big-to-fail banks render the first best in a two-tier architecture unattainable, strengthening the case for CBDC. We also show that, unlike in the single-tier system, optimal policy in the two-tier system with its distortions requires fiscal resources or regulation to correct them. Deadweight losses of taxation or regulation therefore add to the social costs of the two-tier, but not the single-tier payment system.

Our calibrations robustly indicate a cost advantage of CBDC when we adopt this broader perspective. Intuitively, the technological advantage of fractional reserve banking over narrow banking and the social costs of pass-through funding are minor compared to the excess burden of measures to address frictions in the banking sector. Optimally, this translates into a dominant share of CBDC in real balances. A small CBDC share only results if pass-through funding is necessary to stabilize bank lending and very costly and if deposits and CBDC are close substitutes.

We also analyze the case for noncirculating CBDC. When deposits are the efficient retail means of payment but a deposit tax/subsidy to steer banks' price setting is not admissible, a noncirculating CBDC with a suitably chosen interest rate target can alternatively serve to discipline banks (Andolfatto, 2021). We show that this mechanism only operates when the Ramsey policy subsidizes deposits. When it taxes deposits (because strong externalities require high interest on reserves), a CBDC interest rate target cannot replace the missing instrument because it cannot force banks to lower the deposit rate.[2] We also show that even if a deposit subsidy is admissible, a noncirculating CBDC can usefully substitute for it if the subsidy requires tax funding that is distortionary.

A recurrent finding in our analysis is that the two central bank liabilities, reserves and CBDC, should pay different interest rates. Intuitively, the spread on a circulating means

[2] In Andolfatto (2021), no reserves or externalities are present.

of payment should reflect social costs and externalities, which differ between reserves and CBDC. And the interest rate on a noncirculating CBDC disciplining banks should reflect the central bank's target for the deposit rate, which reflects factors unrelated to the social costs of reserves. Implementing the optimal policy therefore requires the government to price discriminate between wholesale and retail holders of central bank liabilities.

Related Literature The paper relates to the literature on money multipliers and the two-tier monetary system with inside vs. outside money (Phillips, 1920; Gurley and Shaw, 1960). Tobin (1963; 1969; 1985) discusses the fractional reserve banking system and proposes a precursor to CBDC. Benes and Kumhof (2012) simulate fractional reserve banking in a DSGE model, and Brunnermeier and Niepelt (2019) and Faure and Gersbach (2018) compare allocations with and without private money creation. Chari and Phelan (2014) emphasize negative externalities of fractional reserve banking when central bank money is scarce while in Taudien (2020), inside money productively lowers financing costs. Our micro foundation for the role of reserves emphasizes fire sales (Shleifer and Vishny, 1992; Stein, 2012) and frictional interbank markets (Bianchi and Bigio, 2022). In Kiyotaki and Moore (2019) a liquid security, like reserves, relaxes resalability constraints, and in Parlour et al. (2022) a transfer of bank deposits generates a costly "liquidity externality."

In Keister and Sanches (2023), the central bank injects CBDC by transfer, promoting exchange but crowding out deposits and deposit-funded investment. In Barrdear and Kumhof (2022), the central bank issues CBDC in exchange for government bonds and this curtails CBDC-deposit substitution. Böser and Gersbach (2020) and Williamson (2022) analyze the role of central bank collateral in the CBDC context, and Piazzesi and Schneider (2021) focus on effects of CBDC on banks' contingent liquidity lines. Schilling et al. (2020) analyze conflicts between efficiency, run risk and price stability in monetary economies including those with CBDC.

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.[3] Chiu et al. (2023) quantitatively assess the implications of CBDC in a framework that combines elements of Andolfatto (2021) and Keister and Sanches (2023). We show that CBDC may help discipline banks when deposit rates are inefficiently low but not otherwise. We also show that scarce public funds render a disciplining, noncirculating CBDC preferable to corrective subsidies.

[3] See Garratt and van Oordt (2021) for disciplining effects of CBDC in the privacy context.

Our analysis complements large literatures building on "New Keynesian" and "New Monetarist" frameworks (see, for example, Galí, 2015; Rocheteau and Nosal, 2017). Unlike the former, we emphasize the role of money as means of payment (Tobin, 1969) and abstract from nominal rigidities; compared to the latter, we emphasize general equilibrium implications rather than the market micro structure underlying the supply and demand for liquidity.

Structure of the Paper Section 2 presents the monetary economy. In section 3 we characterize equilibrium and the monetary policy transmission. Section 4 contains the normative analysis, and section 5 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.

We build on the standard neo-classical growth or real business cycle model augmented with household preferences for liquidity, as in Sidrauski (1967). We introduce three elements into this environment. First, a banking sector. Second, multiple means of payment. At the retail level, bank issued deposits and central-bank issued CBDC circulate. At the wholesale level, central-bank issued reserves serve as means of payment for banks. (We discuss the role of cash below.) Finally, we introduce costs of operating the payment system[4] and engaging in liquidity transformation. These three elements represent key features of the contemporary monetary architecture, in which banks settle client payments using reserves.

The analysis does not impose restrictions on the sources of aggregate risk;[5] all parameters or functions indexed by time may also depend on histories. To streamline notation we only let a few parameters and functions explicitly depend on time and histories.

[4]The costs might also reflect the management of assets backing the payment instruments.

[5]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

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

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

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

Here, c_t denotes household consumption at date t; z_{t+1} denotes "effective real balances" carried from t into $t+1$; $k_{t+1}, m_{t+1}, n_{t+1}$ denote capital, CBDC or "money," and bank deposits carried from t into $t+1$, respectively; R_t^k, R_t^m, R_t^n denote gross rates of return on capital, CBDC and deposits, respectively; and w_t, π_t, τ_t denote wage income, profit income and taxes, respectively.[6]

Effective real balances are a CES composite of money and bank deposits,

$$z_{t+1} \equiv z(m_{t+1}, n_{t+1}; \tilde{\lambda}_t) \equiv \frac{1}{1-\tilde{\lambda}_t} \left(\tilde{\lambda}_t m_{t+1}^{\frac{\eta-1}{\eta}} + (1-\tilde{\lambda}_t) n_{t+1}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}},$$

where $\tilde{\lambda}_t \in (0,1)$ indexes liquidity benefits of money relative to deposits, reflecting factors such as privacy or convenience of use. Parameter $\eta > 1$ represents the elasticity of substitution. Since the elasticity exceeds unity real balances are strictly positive even if only deposits circulate—the model also speaks to a pre-CBDC world. As $\eta \to \infty$ money and deposits become perfectly substitutable and real balances converge to

$$\lim_{\eta\to\infty} z(m_{t+1}, n_{t+1}; \tilde{\lambda}_t) = \lambda_t m_{t+1} + n_{t+1},$$

where $\lambda_t \equiv \tilde{\lambda}_t/(1-\tilde{\lambda}_t) > 0$. A large elasticity η appears plausible; many central banks considering the introduction of CBDC anticipate technological solutions that meet the same regulatory standards as deposit accounts and offer similar types of user experience as deposit based digital payments.[7]

The felicity function u is increasing, strictly concave and satisfies Inada conditions, and the discount factor $\beta \in (0,1)$. As is well known, the "money in the utility function"

[6] Without relevant loss of generality we abstract from preferences for leisure; in the discussion paper we consider the case with elastic labor supply (Niepelt, 2022). Households also hold bank equity; we discuss the implications when analyzing the bank problem.

[7] See for example Auer et al. (2020). Nagel (2016), among others, finds a very high elasticity of substitution between deposits and other liquid assets.

specification can represent a broad set of monetary frictions and flexibly generates a demand for liquidity.[8] Other specifications of money demand would produce similar results. What matters is not why households value liquidity but that they do; valued liquidity services imply that the Modigliani and Miller (1958) theorem does not apply for issuers of real balances.

Equation (1) represents the household budget constraint. The household invests in capital as well as real balances; pays for consumption and taxes; and funds these outlays with wage income, distributed profits and the gross returns on capital, money and deposits. The gross rates of return on means of payment reflect short-term nominal interest rates and inflation. To keep the notation for spreads transparent (see below) we assume that R_t^m and R_t^n are risk-free (inflation risk is negligible). The gross rate of return on capital, R_t^k, may be risky.[9]

We abstract from cash and government bond 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. Letting households hold government bonds would be inconsequential too. We discuss this in detail in appendix A.

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

$$\begin{aligned} u_c(c_t, z_{t+1}) &= \beta \mathbb{E}_t[R_{t+1}^k u_c(c_{t+1}, z_{t+2})], \\ u_c(c_t, z_{t+1}) &\geq \beta R_{t+1}^m \mathbb{E}_t[u_c(c_{t+1}, z_{t+2})] + u_z(c_t, z_{t+1}) z_m(m_{t+1}, n_{t+1}; \tilde{\lambda}_t), \quad m_{t+1} \geq 0, \\ u_c(c_t, z_{t+1}) &\geq \beta R_{t+1}^n \mathbb{E}_t[u_c(c_{t+1}, z_{t+2})] + u_z(c_t, z_{t+1}) z_n(m_{t+1}, n_{t+1}; \tilde{\lambda}_t), \quad n_{t+1} \geq 0, \end{aligned} \tag{2}$$

respectively.[10] The weak inequalities hold with equality when m_{t+1} or n_{t+1} are interior.

To express the Euler equations more compactly, define the risk-free interest rate as

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

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

[8]One such friction is a "shopping time" friction, which renders real balances helpful to economize on time spent shopping (McCallum and Goodfriend, 1987; Croushore, 1993).

[9]Variables c_t, k_{t+1}, m_{t+1}, n_{t+1}, z_{t+1}, R_t^k, R_{t+1}^m, R_{t+1}^n, w_t, π_t, τ_t and parameter $\tilde{\lambda}_t$ are measurable with respect to information available at date t.

[10]We indicate partial derivatives by subscripts.

condition reads

$$\frac{u_z(c_t, z_{t+1}) z_i(m_{t+1}, n_{t+1}; \tilde{\lambda}_t)}{u_c(c_t, z_{t+1})} = \chi^i_{t+1}, \tag{3}$$

where $\chi^i_{t+1} \equiv 1 - R^i_{t+1}/R^f_{t+1}$ denotes the liquidity premium on payment instrument i. Equivalently, $-\chi^i_{t+1}$ equals the spread on payment instrument i compared with a risk-free bond that does not provide liquidity services. According to equation (3), instrument i commands a premium and has a negative spread when it provides liquidity services. When the household holds both payment instruments, then

$$\frac{z_m(m_{t+1}, n_{t+1}; \tilde{\lambda}_t)}{z_n(m_{t+1}, n_{t+1}; \tilde{\lambda}_t)} = \lambda_t \left(\frac{m_{t+1}}{n_{t+1}}\right)^{-\frac{1}{\eta}} = \frac{\chi^m_{t+1}}{\chi^n_{t+1}} \tag{4}$$

and the premium on money exceeds the one on deposits if money is sufficiently more liquid and/or scarce, and vice versa.

2.2 Banks

Banks issue deposits and equity to fund operations as well as 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 in appendix A.

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 hold claims on 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 for introducing CBDC. The 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; we adopt the monopsony assumption for convenience. Alternatively, we could assume that the banking sector is monopsonistically competitive and 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.[11] In either case, our central results would change minimally because our analysis concerns the substitutability

[11]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.

between deposits and money, not the substitutability between different types of deposits and the preferences or market structure underlying it.

On the asset side of their balance sheets banks are price takers. This helps focus on the key questions of interest which concern bank funding and liquidity provision. 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). In the normative analysis in section 4, we allow for frictions on the lending side when assessing the robustness of our results.

Per unit of deposit funding, banks require resources $\nu > 0$ to make payments on behalf of their deposit customers. To introduce a role for reserves, r_{t+1}, we assume that liquidity transformation also requires resources; larger reserve holdings relative to deposits reduce the extent of liquidity transformation and therefore lower these costs.[12] This liquidity centric narrative can be micro founded and is consistent with indirect evidence on private benefits of reserve holdings.[13] One specific micro foundation we develop in appendix B 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 it must sell capital. But asymmetric information or related frictions depress the price of capital when other institutions find themselves in the same situation. Reserve holdings thus generate a positive pecuniary externality and reduce a bank's need to engage in costly measures to make up for liquidity shortfalls.

Other micro foundations have similar implications as we discuss in appendix B. We focus on frictions in interbank markets, which force banks with reserve shortfalls to search for a match or to borrow even more expensively and subject to "stigma" at the central bank discount window (Bianchi and Bigio, 2022). This gives rise to the same key properties we impose: Liquidity transformation is costly for banks because it requires them to engage in what we call "liquidity substitution," and liquidity substitution is costlier when the bank itself or its peers hold fewer reserves. As a consequence, banks do hold reserves but not necessarily the efficient amount.

Formally, we assume that liquidity substitution costs per deposit, $\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1})$, are strictly decreasing in the bank's reserves-to-deposits ratio, $\zeta_{t+1} \equiv r_{t+1}/n_{t+1}$, as well as the

[12] A minimum reserves requirement would act similarly as costs of liquidity transformation. An optimum minimum reserves requirement could place banks in a corner while being optimal for the banking sector as a whole, due to externalities of the kind we consider.

[13] See for example Bianchi and Bigio (2022), Ozdenoren et al. (2021) and Van den Heuvel (2022) for the former and Afonso, Duffie, Rigon and Shin (2022), Afonso, Giannone, La Spada and Williams (2022) and Bianchi and Bigio (2022) for the latter. Reserve holdings are also consistent with narratives that emphasize the costs of managing bank assets (Cúrdia and Woodford, 2009).

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—banks are "narrow banks" and deposits amount to "synthetic CBDC"—then liquidity substitution is not needed, $\omega_t(1,1) = 0 = \omega_{\zeta,t}(1,1) + \omega_{\bar{\zeta},t}(1,1)$. Lower reserve holdings generate liquidity substitution costs. We assume that these costs are sufficiently high such that active banks always find it profitable to hold some reserves; that is, in any equilibrium with deposits, ζ_{t+1} is strictly positive.[14] We focus on symmetric equilibria in which $\zeta_{t+1} = \bar{\zeta}_{t+1}$ and assume that private and social benefits of reserves are decreasing.[15] This implies that $\omega_t(\zeta_{t+1}, \zeta_{t+1})$ is convex.

Each regional bank is long-lived but its program consists of a sequence of static problems, which involve a choice of deposit issuance and reserve holdings that also determines capital investment and equity issuance. Formally, the date-t program of a bank reads[16]

$$\begin{aligned}
\max_{n_{t+1}, r_{t+1}} \quad & \pi_t^{b1} + \mathbb{E}_t[\mathrm{sdf}_{t+1}\ \pi_{t+1}^{b2}] \\
\text{s.t.} \quad & \pi_t^{b1} = -n_{t+1}(\nu + \omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1}) - \theta_t), && (5) \\
& \pi_{t+1}^{b2} = (n_{t+1} - r_{t+1})R_{t+1}^k + r_{t+1}R_{t+1}^r - n_{t+1}R_{t+1}^n, && (6) \\
& R_{t+1}^n \text{ reflects deposit funding schedule,} \\
& n_{t+1} \geq r_{t+1} \geq 0,
\end{aligned}$$

where π_t^{b1} and π_{t+1}^{b2} denote cash flows 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, which are financed by equity issuance. The second constraint relates the return on equity to the returns on capital investment, $n_{t+1} - r_{t+1}$, reserve holdings and deposits.[17] The 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 an operating bank's gross rate of return on equity, $-\pi_{t+1}^{b2}/\pi_t^{b1}$, is increasing in the liquidity premium on deposits and decreasing in the premium on reserves. While this rate of return exceeds the required rate of return, any equity funds in excess of the

[14] This is guaranteed by letting $\omega_t(0,0)$ be sufficiently high.

[15] Formally, we assume that $\omega_{\zeta\zeta,t} + \omega_{\zeta\bar{\zeta},t} > 0$ and $\omega_{\zeta\zeta,t} + \omega_{\zeta\bar{\zeta},t} + \omega_{\bar{\zeta}\zeta,t} + \omega_{\bar{\zeta}\bar{\zeta},t} > 0$ when $\zeta_{t+1} = \bar{\zeta}_{t+1}$.

[16] Variables $r_{t+1}, \zeta_{t+1}, \mathrm{sdf}_t, R_{t+1}^r, \pi_t^{b1}, \pi_t^{b2}, \theta_t$, 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.

[17] When limited liability imposes the constraint $\pi_{t+1}^{b2} \geq 0$, we can interpret (5)–(6) as the cash flows of a bank investor that also insures the bank against negative equity realizations at market prices. Alternatively, one could assume that banks need to issue equity in excess of $-\pi_t^{b1}$ or need to inject "sweat equity" after low realizations of R_{t+1}^k.

optimal $-\pi_t^{b1}$ would be invested in capital. That is, inframarginal bank equity funds profitable monopsony business but additional equity would fund capital investment. The incentive for banks to raise equity for their profitable deposit business, but not more, determines an equilibrium leverage ratio.[18]

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

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

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

$$-(\nu+\omega_t(\cdot)-\theta_t)+\omega_{\zeta,t}(\cdot)\zeta_{t+1}+\chi_{t+1}^n=n_{t+1}R_{t+1}^{n}{}'(n_{t+1})/R_{t+1}^f. \tag{7}$$

The left-hand side of equation (7) represents the marginal profit from deposit issuance, holding the interest rate constant: A marginal unit generates net operating and liquidity substitution costs $\nu+\omega_t(\cdot)-\theta_t$ and increases the liquidity substitution costs for inframarginal units, but it also yields a gain if the deposit liquidity premium is positive, $\chi_{t+1}^n>0$. The right-hand side represents the bank's marginal loss from higher deposit issuance forcing the bank to increase R_{t+1}^n. Equation (7) simplifies to

$$\chi_{t+1}^n-(\nu+\omega_t(\cdot)-\theta_t-\omega_{\zeta,t}(\cdot)\zeta_{t+1})=\frac{1}{\eta_{t+1}^n}\frac{R_{t+1}^n}{R_{t+1}^f},$$

where η_{t+1}^n denotes the elasticity of deposit funding with respect to R_{t+1}^n (see Klein, 1971; Monti, 1972). This elasticity may depend on central bank choices, in particular on whether—and how elastically—the central bank supplies money. We discuss this in detail below.

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

$$-\omega_{\zeta,t}(\zeta_{t+1},\bar{\zeta}_{t+1})=\chi_{t+1}^r, \tag{8}$$

where $\chi_{t+1}^r\equiv 1-R_{t+1}^r/R_{t+1}^f$ denotes the liquidity premium on reserves. Intuitively, the optimal choice of reserves equalizes the private gain from lower liquidity substitution costs and the loss from the excess return on capital over reserves. Since $\zeta_{t+1}=\bar{\zeta}_{t+1}$ in equilibrium, equation (8) implies a unique mapping from the opportunity cost of holding reserves to the equilibrium reserves-to-deposits ratio, which we write as $\zeta_{t+1}=\omega_{\zeta,t}^{-1}(-\chi_{t+1}^r)$.[19] Re-

[18] Deposit relative to equity funding equals $-n_{t+1}/\pi_t^{b1}=1/(\nu+\omega_t(\cdot)-\theta_t)$.

[19] By the mean value theorem, for any $\varepsilon>0$ there exists a $\iota\in(0,\varepsilon)$ such that $\omega_{\zeta,t}(\zeta+\varepsilon,\zeta+\varepsilon)=\omega_{\zeta,t}(\zeta,\zeta)+(\omega_{\zeta\zeta,t}(\zeta+\iota,\zeta+\iota)+\omega_{\zeta\bar{\zeta},t}(\zeta+\iota,\zeta+\iota))\varepsilon$. Function $\omega_{\zeta,t}(\zeta,\zeta)$ thus is monotonically increasing

alistically, our framework therefore implies substitution effects on the asset side of bank balance sheets in response to changes in the spread on reserves.

Combining equations (7) and (8) for an active bank implies

$$\chi^n_{t+1} - \big(\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t\big) = \frac{1}{\eta^n_{t+1}} \frac{R^n_{t+1}}{R^f_{t+1}}, \tag{9}$$

where we define

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

Function $\tilde{\omega}_t$ summarizes how the spread on reserves affects marginal liquidity substitution costs when banks choose reserves privately optimally. Our assumption about function ω_t implies that $\tilde{\omega}_t$ is strictly increasing in χ^r_{t+1}.[20] Condition (9) relates the deposit spread (represented by χ^n_{t+1} and R^n_{t+1}/R^f_{t+1}) and thus balance sheet length to policy (reserve spread and deposit subsidy) and the elasticity of deposit funding.

According to equation (9), the deposit rate reflects the reserve rate as well as bank leverage; ceteris paribus, it co-moves with the risk-free interest rate but bank market power (small η^n_{t+1}) attenuates this effect.[21] This is consistent with theoretical and empirical evidence 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 (see section 3).

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 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

in ζ and therefore invertible (see footnote 15).

[20]Define $\hat{\omega}_t(\zeta) \equiv \omega_t(\zeta,\zeta) - \omega_{\zeta,t}(\zeta,\zeta)\zeta$ and note that $\tilde{\omega}_t$ equals $\hat{\omega}_t$ subject to $\zeta_{t+1} = \omega^{-1}_{\zeta,t}(-\chi^r_{t+1})$. Differentiating yields $\hat{\omega}'_t = \omega_{\zeta,t} + \omega_{\bar{\zeta},t} - \omega_{\zeta,t} - \zeta(\omega_{\zeta\zeta,t} + \omega_{\zeta\bar{\zeta},t}) < 0$ (see footnote 15). Since ζ_{t+1} is strictly decreasing in χ^r_{t+1} we conclude that $\tilde{\omega}_t$ is strictly increasing in χ^r_{t+1}.

[21]An alternative representation of condition (9) is $R^f_{t+1}(1 - \nu - \tilde{\omega}_t(-\chi^r_{t+1}) + \theta_t) = R^n_{t+1}(1/\eta^n_{t+1} + 1)$. The inverse leverage ratio equals $\nu + \omega_t(\cdot) - \theta_t$.

solves[22]

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

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 $\mu > 0$, and the unit resource costs of managing reserve based payments among banks equal $\rho > 0$. Accordingly, the government budget constraint reads[23]

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

The government's seignorage revenue equals $m_{t+1}\chi_{t+1}^m + r_{t+1}\chi_{t+1}^r$, which compares with $n_{t+1}(\chi_{t+1}^n - \zeta_{t+1}\chi_{t+1}^r)$ in the private sector.

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; this reduces the central bank's net worth but does not change the allocation because transfers and the taxes funding them are nondistorting and households homogeneous.

[22] Variables π_t^f, κ_{t+1} and ℓ_t are measurable with respect to information available at date t.
[23] Variable k_{t+1}^g is measurable with respect to information available at date t.

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 imply

$$\ell_t = 1, \quad \kappa_t = k_t + k_t^g + n_t - r_t, \quad \pi_t = \pi_t^{b1} + \pi_t^{b2} + \pi_t^f. \tag{14}$$

2.6 Resource Constraint

Consistent with Walras' law, equations (1), (5), (6), (10), (13), and (14) yield the resource constraint or goods market clearing condition

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

The three right-most terms 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

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

2.7 Policy and Equilibrium

Let $\xi_{t+1} \equiv m_{t+1}/n_{t+1}$ denote the ratio of money to deposits when $n_{t+1} > 0$. 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 ratio, respectively.

An *equilibrium* conditional on $\mathcal{P}$ and an initial state consists of

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

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 analyze how the government can optimally exploit instrument redundancy.

2.8 Functional Form Assumptions

Existence and uniqueness of equilibrium in models with real balances in the utility function may require conditions on primitives (see, for example, Walsh, 2017). Our analysis applies for general functional forms, u, that satisfy these conditions. When results require more structure, we selectively impose the following functional form assumptions:

Assumption 1. Preferences satisfy

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

where $\vartheta, \psi \in (0,1)$ and $\sigma > 0, \neq 1$.

Under assumption 1, the elasticity of substitution between c_t and z_{t+1} equals $\psi^{-1} > 1$.

Assumption 2. Bank resource costs of 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 $\varphi > 1, \bar{\varphi} > 0$ and ϕ_t is sufficiently large for active banks to choose $\zeta_{t+1} > 0$.

In appendix B we discuss two micro foundations for this specification, one related to fire sales and the other to frictional interbank markets.

3 General Equilibrium

3.1 Model Solution

The equilibrium conditions reduce to a few core equations. The general equilibrium Euler equation, which combines conditions (2), (11) and (14), and the resource constraint (15) characterize the equilibrium allocation conditional on means of payment:[24]

$$
\begin{aligned}
u_c(c_t, z_{t+1}) &= \beta \mathbb{E}_t[(1-\delta + f_{\kappa,t+1}(\kappa_{t+1}, 1))u_c(c_{t+1}, z_{t+2})], \\
\kappa_{t+1} &= f_t(\kappa_t, 1) + \kappa_t(1-\delta) - c_t - m_{t+1}\mu - n_{t+1}(\nu + \omega_t(\zeta_{t+1}, \zeta_{t+1})) - r_{t+1}\rho.
\end{aligned}
$$

Except for the presence of means of payment, this system parallels the equilibrium conditions in the neoclassical growth model; except for the decomposition of real balances into money and deposits and the payment operations and liquidity substitution costs in the resource constraint, it resembles the equilibrium conditions in a Sidrauski (1967) type model.

The remaining equilibrium conditions (3), (4), (7) and (8) determine the means-of-payment side of the model conditional on consumption and policy: The former two conditions characterize liquidity demand for m_{t+1} and n_{t+1} depending on c_t, χ^m_{t+1} and χ^n_{t+1}, and the latter two conditions pin down χ^n_{t+1} and ζ_{t+1} conditional on the deposit funding schedule and $\mathcal{P}$. That is, conditional on $\mathcal{P}$ the four conditions map c_t into $(m_{t+1}, n_{t+1}, \zeta_{t+1}, \chi^n_{t+1})$ and thus (z_{t+1}, r_{t+1}).

3.2 Monetary Policy Transmission

The transmission mechanism of monetary policy operates through the cost structure of banks and their price setting.[25] Impose for concreteness assumption 1 such that households' elasticity of substitution between consumption and real balances equals ψ^{-1}. Suppose in addition that the central bank targets $\xi_{t+1} \equiv m_{t+1}/n_{t+1}$, for instance because it sets $m_{t+1} = 0$, such that the elasticity of deposit funding with respect to R^n_{t+1} in equation (9) satisfies $1/\eta^n_{t+1} = \psi \chi^n_{t+1} R^f_{t+1}/R^n_{t+1}$ (see appendix C). Optimality condition (9) then reads

$$
\chi^n_{t+1} = \frac{\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t}{1-\psi}; \tag{9'}
$$

[24] The equilibrium conditions also include the restrictions that asset holdings be nonnegative. The budget constraints (1), (5), (6), (10), and (13) determine the ownership structure of capital and profits. Equation (12) determines labor income.

[25] See Niepelt (2022) for computed impulse responses.

that is, banks charge a constant markup over costs. Both higher deposit subsidies and higher interest on reserves raise the equilibrium deposit rate and real balances.[26]

Higher interest on reserves thus lengthens bank balance sheets. The strength of this effect increases with the extent to which higher interest on reserves lowers $\tilde{\omega}_t$ and thus the numerator in equation (9'); the cost reduction translates into a lower deposit spread (markup factor $(1-\psi)^{-1}$) and more deposit funding (elasticity ψ^{-1}). On the other hand, higher interest on reserves induces substitution effects on the asset side of bank balance sheets as we saw in section 2. The strength of this second effect depends on the elasticity of reserve holdings. Depending on the relative magnitude of the two effects, higher interest on reserves may drive bank loans up or down.

Suppose next that the central bank targets χ^m_{t+1} rather than ξ_{t+1}. The elasticity of deposit funding in equation (9) then depends on the composition of real balances. An instructive special case arises when money and deposits are perfect substitutes ($\eta \to \infty$) such that households generically only use the cheaper payment instrument (see condition (4)). There are two possibilities: Either the central bank sets χ^m_{t+1} higher than λ_t times the "monopsony premium," defined as the equilibrium deposit premium χ^n_{t+1} when the central bank does not issue money. This renders money unattractive for households, confronts banks with the same elasticity of deposit funding as above, and thus gives rise to the same monetary transmission mechanism.

Or, the central bank targets χ^m_{t+1} below λ_t times the monopsony premium. Banks then have a choice between lowering the deposit spread to the competitive level χ^m_{t+1}/λ_t and pricing the central bank out of the market, or not lowering the spread 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.[27] By setting χ^m_{t+1}, the central bank directly controls χ^n_{t+1} in this case.

Suppose finally that the central bank targets $m_{t+1} > 0$ rather than χ^m_{t+1} or ξ_{t+1}. The elasticity of deposit funding in equation (9) then again depends on the composition of real balances. But the quantity competition between banks and the central bank is less intense than the price competition when the central bank targets χ^m_{t+1}.

[26]Recall that $\tilde{\omega}_t$ increases in χ^r_{t+1}. The interest rate instrument is fiscally cheaper, see Niepelt (2022).

[27]If the reverse inequality holds, banks would make losses when trying to compete with the central bank. In equilibrium, n_{t+1} then equals zero.

4 Optimality

Should CBDC be introduced? And how should the central bank manage its liabilities—reserves and CBDC—to optimally conduct monetary policy? To address these questions, we start by characterizing and quantifying the choices of a social planner that is constrained by the production and payment technologies but faces no instrument admissibility restrictions. In a second step, we analyze whether and how a Ramsey government can implement the planner allocation and asset holdings and whether admissibility restrictions on policy instruments bind. Finally, we study the implications of a range of added frictions related to the bank lending channel and costly central bank loans to banks; tax distortions; and too-big-to-fail banks, and we discuss optimal interest rate policies. With each added friction, we need to calibrate additional (conceptually unrelated) parameters to quantify our answers to the posed questions.

4.1 Social Planner Allocation

The social planner solves

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

The optimality conditions for capital and consumption yield the same Euler equation as in equilibrium, see subsection 3.1. Accordingly, our analysis focuses on the planner's choice of money, deposits and reserves.

Start with the latter. To maximize the efficiency of liquidity transformation the planner issues reserves until marginal benefits due to lower liquidity substitution costs equal marginal costs,

$$\left.\begin{array}{rl} \omega_{\zeta,t}(\zeta_{t+1}, \zeta_{t+1}) + \omega_{\bar{\zeta},t}(\zeta_{t+1}, \zeta_{t+1}) + \rho = 0 & \text{if} \quad n_{t+1} > 0 \\ r_{t+1} = 0 & \text{otherwise} \end{array}\right\}. \tag{SP-1}$$

When reserves generate externalities, $\omega_{\bar{\zeta},t} \neq 0$, the planner takes this into account—unlike a bank, which only internalizes private benefits, see equation (8). The shape of the ω_t function implies a unique mapping from ρ into the planner's optimal ζ_{t+1} choice;[28] we

[28] By the mean value theorem, for any $\varepsilon > 0$ there exists a $\iota \in (0, \varepsilon)$ such that $\omega_{\zeta,t}(\zeta+\varepsilon, \zeta+\varepsilon) + \omega_{\bar{\zeta},t}(\zeta+$

denote this choice by $\zeta^{\star}_{t+1}$. Our assumption about ω_t and the fact that $\rho > 0$ imply that $\zeta^{\star}_{t+1} < 1$.

Turning to the choice of money, the optimality condition

$$u_z(c_t, z_{t+1}) z_m(m_{t+1}, n_{t+1}; \tilde{\lambda}_t) \leq u_c(c_t, z_{t+1})\mu, \;\; m_{t+1} \geq 0, \tag{SP-2}$$

reflects the standard Friedman (1969) logic according to which the planner provides liquidity until social marginal benefits and costs are equalized. In its common form, the Friedman (1969) rule stipulates zero marginal costs and thus calls for satiation; here, social costs are strictly positive because payment operations require resources, $\mu > 0$. A parallel logic applies for deposits, which generate social costs because of liquidity substitution as well as deposit and reserve operations:

$$u_z(c_t, z_{t+1}) z_n(m_{t+1}, n_{t+1}; \tilde{\lambda}_t) \leq u_c(c_t, z_{t+1})(\nu + \omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1}) + \zeta^{\star}_{t+1}\rho), \;\; n_{t+1} \geq 0. \tag{SP-3}$$

When the planner uses both money and deposits, then (SP-2) and (SP-3) imply

$$\left(\frac{m_{t+1}}{n_{t+1}}\right)^{-\frac{1}{\eta}} = \frac{\mu/\lambda_t}{\nu + \omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1}) + \zeta^{\star}_{t+1}\rho}, \tag{SP-4}$$

the counterpart of equilibrium condition (4). We have the following result:

Proposition 1. The social planner minimizes the unit social costs of effective real balances. When $\eta \to \infty$, the planner generically provides a single retail means of payment, namely money (deposits) if and only if $\mu/\lambda_t < (>) \nu + \omega_t(\zeta^{\star}_{t+1}, \zeta^{\star}_{t+1}) + \zeta^{\star}_{t+1}\rho$; otherwise, both money and deposits circulate according to condition (SP-4). Optimal quantities satisfy conditions (SP-1)–(SP-3).

Proof. The dual of the planner program implies the first statement. For the second statement, note that the Inada condition rules out $z_{t+1} = 0$, so at least one retail means of payment is used. Consider first $\eta \to \infty$ and suppose $m_{t+1}, n_{t+1} > 0$ such that the left-hand side of (SP-4) equals unity. Generically, the right-hand side of (SP-4) differs from unity and this yields a contradiction. Consider next $\eta < \infty$ and suppose $m_{t+1} = 0 < n_{t+1}$. Since $\lim_{m \downarrow 0} z_m(m, n; \tilde{\lambda}) \to \infty$, (SP-2) yields a contradiction. A parallel argument rules out $n_{t+1} = 0 < m_{t+1}$ if $\eta < \infty$. □

Recall that $n^{\star}_{t+1} > 0$ implies $\zeta^{\star}_{t+1} < 1$: A planner that opts for deposits backs them

$\varepsilon, \zeta + \varepsilon) = \omega_{\zeta,t}(\zeta, \zeta) + \omega_{\bar{\zeta},t}(\zeta, \zeta) + (\omega_{\zeta\zeta,t}(\zeta + \iota, \zeta + \iota) + 2\omega_{\zeta\bar{\zeta},t}(\zeta + \iota, \zeta + \iota) + \omega_{\bar{\zeta}\bar{\zeta},t}(\zeta + \iota, \zeta + \iota))\varepsilon$. Function $\omega_{\zeta,t}(\zeta, \zeta) + \omega_{\bar{\zeta},t}(\zeta, \zeta)$ thus is monotonically increasing in ζ and therefore invertible (see footnote 15).

only partly with reserves because liquidity substitution is costless at the margin when $\zeta_{t+1} = 1$. This has an important implication:

Lemma 1. Suppose that money and deposits are perfect substitutes, $\eta \to \infty$. If narrow bank deposits provide liquidity more efficiently than money, $\nu + \rho \leq \mu/\lambda_t$, the planner does not issue money.

Proof. Deposits subject to optimal reserve holdings provide liquidity at unit cost $\nu + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \zeta^\star_{t+1}\rho = \nu + \rho + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) - (1 - \zeta^\star_{t+1})\rho < \nu + \rho$, where the inequality follows from the convexity of ω_t and (SP-1). Since $\nu + \rho \leq \mu/\lambda_t$ by assumption, proposition 1 implies the result. □

Lemma 1 gives a necessary condition for the planner to issue money when money and deposits are perfect substitutes: Liquidity provision using money must be at least as efficient as a narrow banking arrangement. This suggests narrow banking as a natural benchmark to assess the efficiency of liquidity provision. We use this benchmark in the quantitative analysis.

Quantitative Implications We impose assumption 2 and use monthly return and balance sheet data over the period 2010–2019 as well as information about resource costs of payment systems in the private and public sector and about the effects of bank liquidity on financial stability, among other sources, to calibrate ν, ρ and ω. Appendix D contains detailed discussions and table 1 summarizes the calibration and our sources.

We observe an average annual liquidity premium on reserves, χ^r, of roughly 50 basis points (bp) and an average reserves-to-deposits ratio, ζ, of around 19.5%. We rely on bank data to gauge noninterest expenses associated with deposit-taking and on central bank studies to gain a comprehensive perspective on the costs of payment systems. Focusing on the variable macroeconomic costs, we calibrate ν to roughly 80 bp annually.[29] Based on Fedwire cost data and international evidence on central bank shares of payment operation costs (net of cash handling), we calibrate ρ to about 2 bp annually. The fact that measured χ^r is much larger than measured ρ suggests that χ^r is suboptimally high and ζ suboptimally low (see the Ramsey analysis below). It does not indicate problems with the calibration.[30]

[29] We discuss the role of fixed costs below.

[30] Any ω function with weakly positive externalities implies $\chi^r \leq \rho$ under the optimal policy (see the Ramsey analysis below). While the model explains the observed ζ conditional on the observed χ^r it does not rationalize the choice of χ^r.

To calibrate ω, we use three model implied moments: The bank's demand for reserves conditional on the observed χ^r, which we relate to the measured ζ; the elasticity of banks' reserves demand, which we relate to its empirical counterpart; and the effect of changes in reserve holdings on liquidity substitution, which we relate to an empirical measure of bank vulnerability. This and other information implies that ω is convex with an intercept of roughly 50 bp. See appendix D for detailed discussions.

	Range	Sources
Liquidity premium on reserves; reserves-to-deposits ratio		
χ^r	$[0.3976, 0.5964] \cdot 10^{-2}$	FRED, van Binsbergen et al. (2022)
ζ	$[0.1556, 0.2334]$	
Deposit operating costs; reserve operating costs		
ν	$[0.4867, 1.1933] \cdot 10^{-2}$	Benati (2020), BIS, Danmarks Nationalbank (2018a; 2018b), Federal Reserve Board, Hanson et al. (2015), Junius et al. (2022), Norges Bank (2022), Schmiedel et al. (2012), Sintonen and Takala (2022), Statistics Norway, Stewart et al. (2014), Trebbi and Zhang (2022), Van den Heuvel (2022), Wang et al. (2020)
ρ	$[0.9634, 5.3277] \cdot 10^{-4}$	
Liquidity substitution costs		
ϕ	$[0.2968, 0.8547] \cdot 10^{-2}$	Afonso, Giannone, La Spada and Williams (2022), Begenau and Landvoigt (2022), Bennet and Unal (2015), Bianchi and Bigio (2022), Corbae and D'Erasmo (2021), Duarte and Eisenbach (2021), FRED, Roberts et al. (2023), van Binsbergen et al. (2022)
φ	$[1.0208, 1.9860]$	
$\bar{\varphi}$	$[0.9385, 1.8789]$	

Table 1: Calibration. See the text and appendix D for explanations.

Since we use multiple data sources and allow for measurement error, we obtain multiple calibrations. Rather than averaging or aggregating in other ways, we acknowledge the uncertainty about model parameters and assess the robustness of our findings based on it. When data sources imply a range of possible values for an input in the calibration, we treat that input as uniformly distributed between the minimum and maximum values, see table 1. Drawing repeatedly from the distributions of the inputs, we derive a calibration for each input combination, compute the statistics of interest for each calibration and thus arrive at distributions of the statistics of interest.

Turning to the model implications, suppose first that $\eta \to \infty$ such that the ratio on the right-hand side of condition (SP-4), $(\mu/\lambda)/(\nu + \omega^\star + \zeta^\star \rho)$, determines whether the introduction of CBDC is beneficial or not. We stipulate $\lambda = 1$, the most plausible value and natural benchmark;[31] re-normalizing immediately yields results for different λ values.

Lemma 1 implies that for $\eta \to \infty$, CBDC can only be attractive if it provides liquidity more efficiently than narrow banks. We therefore assess the case for CBDC based on the statistic

$$\alpha^\star \equiv \frac{\nu + \rho}{\nu + \omega(\zeta^\star, \zeta^\star) + \zeta^\star \rho} - 1, \tag{16}$$

which represents the efficiency advantage of CBDC relative to a narrow bank arrangement that is required for the planner to choose CBDC over a two-tier system with optimum reserves-to-deposits ratio. Stated differently, μ must be smaller than $(\nu + \rho)/(1 + \alpha^\star)$ to make CBDC preferable over an optimally managed two-tier system. We also compute a second statistic,

$$\alpha \equiv \frac{\nu + \rho}{\nu + \omega(\zeta, \zeta) + \zeta \rho} - 1, \tag{17}$$

which depends on the actual (measured) reserves-to-deposits ratio, ζ, rather than the higher optimal one.

Figure 1 displays histograms of $2'000'000$ realizations of $\alpha^\star$ (in dark gray) and α (light gray). The histogram of $\alpha^\star$ indicates that a cost advantage of CBDC over narrow banks of less than 76 bp suffices to render CBDC advantageous relative to the best possible two-tier system. The histogram of α illustrates that CBDC could exhibit large cost disadvantages relative to narrow banks (of between 9 and 49%) and still be preferable to the status quo.

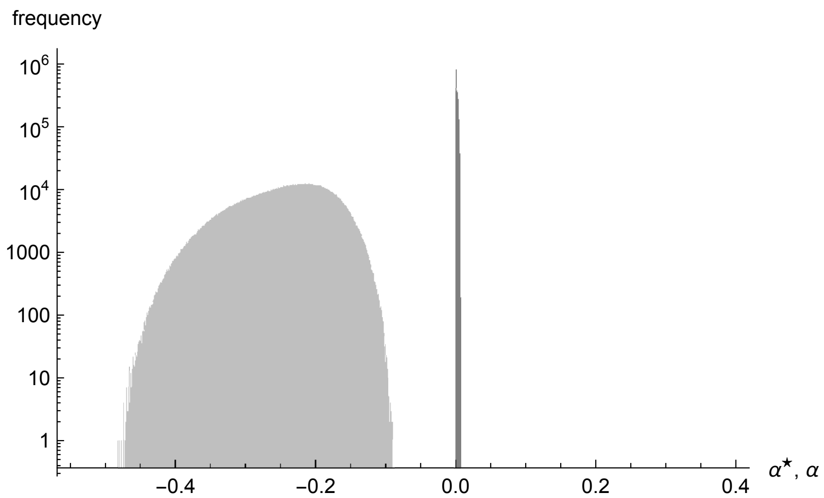

Figure 1: Histograms of $\alpha^\star$ (dark gray) and α (light gray); $2'000'000$ draws each, $1'000$ bins, logarithmic scale.

[31] See the discussion preceding footnote 7.

The large discrepancy between the $\alpha^\star$ and α distributions reflects two factors: First, fractional reserve banking with optimal reserve holdings saves relatively few resources compared with the wastage when reserve holdings are suboptimal ($\omega(\zeta^\star, \zeta^\star) + \zeta^\star\rho < \rho \ll \omega(\zeta, \zeta) + \zeta\rho$). This explains the sign difference between $\alpha^\star$ and α and the lack of symmetry around the origin. Second, the different variances of the two distributions result because draws of $\omega(\zeta, \zeta) + \zeta\rho$ exhibit much larger variation than draws of either ρ or $\omega(\zeta^\star, \zeta^\star) + \zeta^\star\rho$. In conclusion, with $\eta \to \infty$ and $\lambda = 1$ the two-tier system dominates a single-tier system, but only after a rise of the reserves-to-deposits ratio towards its optimal level.

Suppose next that $\eta < \infty$. With a finite elasticity of substitution, the planner issues both deposits and CBDC—the question is how balanced the optimal composition of real balances is.[32] We conservatively assume that CBDC exhibits no efficiency advantage over narrow banks, $\mu = \nu + \rho$. From condition (SP-4) and $\lambda = 1$, the optimal ratio of CBDC to deposits then equals $m^\star/n^\star = (1 + \alpha^\star)^{-\eta}$. When $\eta = 10$ this ratio never falls below 92% in any of the calibration draws, and for $\eta = 50$ the ratio never falls below 68% (averages and medians exceed 91% in both cases). We conclude that a finite elasticity of substitution implies a substantial optimal share of CBDC in real balances even if CBDC payment operations are as costly as narrow banking.

4.2 Ramsey Policy

Unlike the social planner, the Ramsey government controls the allocation and asset holdings only indirectly by choosing an admissible policy $\mathcal{P}$. This raises the question whether the Ramsey policy $\mathcal{P}^\star$ implements the planner allocation and asset holdings and what the policy is. Recall from subsection 4.1 that any equilibrium satisfies the social planner's optimality conditions for capital and consumption as well as the resource constraint. To answer our question, we can therefore focus on the implementability of the first-best quantities of money, deposits and reserves. Accordingly, we analyze the equilibrium conditions that determine m_{t+1}, n_{t+1} and r_{t+1}.

Consider money first. A comparison of conditions (3) and (SP-2) implies the optimal liquidity premium on money

$$\chi_{t+1}^{m\star} \equiv \mu. \tag{RA-1}$$

Intuitively, the optimal liquidity premium reflects social costs; the traditional Friedman rule follows when $\mu = 0$. Note that $\chi_{t+1}^{m\star}$ implements efficient money holdings even if the social planner chooses a corner solution without deposits, which is a possibility when

[32] Realistically, the issuance of CBDC would involve fixed costs, which we abstract from. As a consequence, model implied $m^\star/n^\star$ ratios strengthen the case for CBDC only if they are sufficiently high.

$\eta \to \infty$. In this case, households would only hold deposits if $\chi^n_{t+1} \leq \chi^m_{t+1}/\lambda_t = \mu/\lambda_t$ (see equation (4)) but charging such a low deposit premium would generate losses for banks.[33]

Turning to deposits and reserves, consider first a relaxed Ramsey program without the bank's optimality condition for deposits, equation (7), as a constraint. In this relaxed program, the government implements the first best by ensuring that the deposit liquidity premium equals

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

and the liquidity premium on reserves

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

Equation (RA-2) again follows from comparing the equilibrium conditon (3) and the planner's optimality condition (SP-3). Similarly, conditions (8) and (SP-1) imply equation (RA-3). Intuitively, the optimal deposit premium reflects the social costs of deposits, and the optimal premium on reserves induces banks to choose the first-best reserves-to-deposits ratio although private and social benefits differ. We conclude that in the relaxed program the Ramsey government implements first-best asset holdings.

The actual Ramsey program is subject to the additional equilibrium constraint (7) or (9). But this additional constraint does not materially 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 (9), the optimal subsidy 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 $R^{n\star}_{t+1} \equiv R^{f\star}_{t+1}(1 - \chi^{n\star}_{t+1})$.[34] Using optimality condition (RA-2) as well as the definition of the $\tilde{\omega}_t$ function and of $\zeta^\star_{t+1}$ in condition (SP-1), this can be expressed as

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

[33] The government could ensure this by setting θ_t low and/or χ^r_{t+1} high.

[34] In general, $\mathcal{P}^\star$ may not be unique nor uniquely implement the first best. When we impose functional form assumptions, this is not an issue.

The optimal subsidy in (RA-4) has two components, reflecting the two frictions in the banking sector. More market power (a lower elasticity of deposit funding, η^n_{t+1}) requires a higher subsidy as banks must be encouraged to expand their balance sheets. Maybe more surprisingly, a stronger reserves externality demands the opposite (recall that $\omega_{\bar{\zeta},t} < 0$). Intuitively, higher external benefits lead the Ramsey government to increase the interest rate on reserves, and to sterilize the effect of this higher subsidy on the deposit margin it lowers the deposit subsidy. We have the following result:

Proposition 2. The Ramsey policy implements the first best. If $\eta < \infty$ it satisfies conditions (RA-1)–(RA-4); otherwise (when the planner chooses a corner solution) it satisfies the relevant subset of these conditions. Optimal interest rates on money and reserves generically differ. Deposits may optimally be taxed or subsidized.

Proof. Conditions (RA-1) and (RA-3) and the fact that μ is unrelated to $-\omega_{\zeta,t}(\zeta^\star_{t+1}, \zeta^\star_{t+1})$ or ρ imply that optimal interest rates on money and reserves differ. The other results were derived in the text. □

Two factors cause the optimal remuneration on CBDC and reserves to differ. First, spreads should reflect social costs, and the social costs of CBDC based payments, μ, generically differ from those of reserve based payments, ρ. Second, spreads should also reflect Pigouvian subsidies. Since only reserve holdings generate externalities, only reserve spreads should be lowered relative to payment operations costs. We note that differential remuneration of reserves and CBDC 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. Specifically, under assumption 2 the optimal liquidity premium on reserves equals[35]

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

which falls short of the operating costs of reserve based payments if reserves generate externalities ($\bar{\varphi} > 0$). For $\rho \to 0$ reserves optimally exhibit no liquidity premium and the optimal liquidity premium on deposits approaches ν.

Under preference assumption 1 the optimal subsidy is given by

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

[35] See appendix E for derivations.

when the government targets the composition of real balances.[36] The subsidy equals zero when the elasticity of substitution between consumption and real balances approaches infinity ($\psi \to 0$, such that banks lack market power) and reserves do not generate externalities ($\chi^{r\star}_{t+1} = \rho$), or when $\psi > 0$ and $\chi^{r\star}_{t+1} < \rho$ but the two effects cancel. Deposits are taxed when banks have limited market power (small ψ) and reserves generate substantial externalities ($\rho > \chi^{r\star}_{t+1}$).

Quantitative Implications To quantify the Ramsey policy under the assumptions spelled out above, we calibrate the elasticity of substitution between consumption and real balances, ψ^{-1}. We follow Pasqualini (2021) who argues that consistent with findings in Drechsler et al. (2017) or Wang et al. (2020) the markdown in U.S. deposit markets is 1.5, which translates into $\psi = 1/3$ (see appendix C).

On an annual basis, this implies an optimal spread on reserves, $\chi^{r\star}$, between 0.4 and 3.3 bp and an optimal deposit subsidy, $\theta^\star$, between 15 and 40 bp. The implied spread on deposits, $\chi^{n\star}$, lies between 0.50 and 1.24%. That is, reserves should pay nearly the risk-free interest rate; deposits should (still) be subsidized; and deposits should on average pay roughly 86 bp less than the risk-free rate. The optimal spread on CBDC, $\chi^{m\star}$, depends on the social cost parameter μ. When $\mu = \nu + \rho$ then $\chi^{m\star}$ is only slightly larger than $\chi^{n\star}$.

4.3 Instrument Restrictions and Instrument Redundancy

In addition to the spread on reserves—a common monetary policy instrument—the Ramsey policy relies on a—less common—deposit subsidy, θ_t, to shape bank balance sheets.[37] What happens when such an instrument is not admissible?

Equilibrium conditions (3), (4) and (7) make clear that it is generally impossible to implement $m^\star_{t+1}$, $n^\star_{t+1}$ and $r^\star_{t+1}$ when the θ_t instrument is not admissible—the χ^m_{t+1} and χ^r_{t+1} instruments are insufficient to fully align private and social incentives. The case is more promising when money and deposits are perfect substitutes ($\eta \to \infty$) such that the social planner relies on a single retail payment instrument. When the planner only relies on money, the θ_t instrument is not needed and the instrument restriction is irrelevant. When the planner only relies on deposits, incentives must be aligned along the reserve and deposit margins. Setting the reserve spread to the efficient level, $\chi^{r\star}_{t+1}$, achieves the first goal. Attaining the second goal might also be feasible when the central bank targets the premium on money, χ^m_{t+1}, without actually issuing money.

[36] See appendix E for derivations.

[37] The θ_t instrument could also represent regulatory constraints or deposit insurance.

From equation (9), a bank whose deposits are not subsidized or taxed sets the deposit premium equal to the monopsony premium,

$$\chi_{t+1}^n = \nu + \tilde{\omega}_t(-\chi_{t+1}^{r\star}) + \frac{1}{\eta_{t+1}^n} \frac{R_{t+1}^n}{R_{t+1}^f}.$$

When the central bank stands ready to issue money at a premium χ_{t+1}^m higher than λ_t times the monopsony premium, then banks optimally keep χ_{t+1}^n unchanged. But when the government threatens to issue money at a premium lower than λ_t times the monopsony premium, banks have two options: They may raise the deposit rate, or they may exit (see subsection 3.2). Banks choose the former option if this yields nonnegative profits. Setting $\chi_{t+1}^m = \lambda_t \chi_{t+1}^{n\star}$ therefore implements the first best as long as

$$\nu + \tilde{\omega}_t(-\chi_{t+1}^{r\star}) \leq \chi_{t+1}^{n\star} < \nu + \tilde{\omega}_t(-\chi_{t+1}^{r\star}) + \frac{1}{\eta_{t+1}^n} \frac{R_{t+1}^n}{R_{t+1}^f}.$$

Since the monopsony premium exceeds $\chi_{t+1}^{n\star}$ when $\theta_t^\star$ is positive, and vice versa, we have established the following results:

Proposition 3. When $\eta < \infty$, implementing the first best requires the θ_t instrument. When $\eta \to \infty$ and the social planner relies on deposits, targeting χ_{t+1}^m (in addition to χ_{t+1}^r) can substitute for the first-best deposit subsidy, $\theta_t^\star$, but only if $\theta_t^\star \geq 0$.

Intuitively, when $\eta \to \infty$, $n_{t+1}^\star > 0$ and $\theta_t^\star \geq 0$, then a χ_{t+1}^m target works for the same reason as in Andolfatto (2021): By offering money at attractive terms, the central bank drives deposit rates and thus real balances up.[38] When $\theta_t^\star < 0$, in contrast, monetary policy is powerless because it cannot push deposit rates down; the Ramsey policy then implements a second-best outcome, typically by issuing some money and setting $\chi_{t+1}^r \neq \chi_{t+1}^{r\star}$.[39] The case of powerless monetary policy does not arise in Andolfatto (2021) because that model does not feature reserves that generate externalities.

Stepping beyond the model, time to build and fixed costs make it impossible to attain the first best even if $\theta_t^\star \geq 0$ (and $\eta \to \infty$, $n_{t+1}^\star > 0$). If the introduction of CBDC takes time, banks have no incentive to lower the deposit spread before the CBDC infrastructure is in place. If it requires prior investment whose fixed costs exceed the permanent welfare loss due to the missing θ_t instrument, a government's threat to introduce CBDC is not credible. Finally, even if the fixed costs are lower such that the threat is credible, the

[38] Burlon et al. (2022) argue that news about the European Central Bank's digital euro project affected share prices of European banks, especially those more reliant on deposit funding.

[39] See Niepelt (2020a).

allocation is not first best because a social planner would not have paid the fixed costs.

4.4 Lending Frictions

The analysis so far has focused on the liability side of bank and central bank balance sheets. This was warranted by the fact that liabilities generate resource costs and liquidity benefits, and it was based on the assumption that banks and the central bank (and households) are equally well equipped to hold physical capital; this rendered the capital ownership structure irrelevant. One may question the latter assumption and ask whether the implications of the liability and liquidity centric view remain valid when the ownership structure of capital matters, for instance because of a bank lending channel (Bernanke and Blinder, 1988). We address this question next.

Conservatively, we adopt the extreme view that only banks can fund certain investment projects and that these projects generate large welfare benefits. As a consequence, banks' capital exposure must not fall below their exposure prior to the introduction of CBDC. Each unit of deposit funding lost to CBDC must then be replaced by $1 - \zeta_{t+1}$ units of fresh funds for banks.[40] In principle, these fresh funds can come from the central bank, which gains funds from CBDC issuance.[41] We conclude that a bank lending channel or related frictions only undermine the liability and liquidity centric view if the central bank cannot pass its newly acquired CBDC funds through to banks.

Plausibly, pass-through funding is feasible but costly and this modifies the quantitative implications. Possible drivers of pass-through costs include political economy frictions, which make it difficult for the central bank to lengthen its balance sheet and lend to banks without facing costly pressure from interest groups. Costs could also arise as a consequence of information frictions, which render it difficult for the central bank to supply pass-through funding at adequate terms, thereby changing the elasticity of credit. Yet another cause of friction could be the central bank's insistence on scarce collateral for its loans to banks, which contrasts with the willingness of depositors to provide unsecured funding.[42]

[40] Only $1 - \zeta_{t+1}$ units are required because banks reduce reserve holdings in line with deposits.

[41] Equivalence results establish conditions under which the introduction of CBDC coupled with pass-through funding is neutral (Brunnermeier and Niepelt, 2019; Niepelt, 2020a,b). Pass-through funding can occur indirectly, through markets. The direct pass-through scenario we consider therefore implies an upper bound for pass-through costs.

[42] Williamson (2022) argues that a CBDC arrangement could reduce aggregate collateral scarcity.

Quantitative Implications To account for pass-through costs, we modify the $\alpha^\star$ statistic in (16) to

$$\alpha^{\star \mathrm{p}} \equiv \frac{\nu+\rho}{\nu+\omega(\zeta^\star,\zeta^\star)+\zeta^\star\rho-(1-\zeta^\star)o}-1, \tag{18}$$

where o denotes unit pass-through costs. The numerator of the fraction represents the marginal cost of liquidity provision by a narrow bank while the denominator represents marginal costs in the two-tier system subject to optimal reserve holdings and taking pass-through cost savings into account. Parallel to $\alpha^\star$, statistic $\alpha^{\star \mathrm{p}}$ denotes the efficiency advantage of CBDC relative to a narrow bank that is required to make a single-tier system more resource efficient than a two-tier system. Similarly, we modify the α statistic in (17), which is based on actual rather than optimal reserve holdings, to

$$\alpha^{\mathrm{p}} \equiv \frac{\nu+\rho}{\nu+\omega(\zeta,\zeta)+\zeta\rho-(1-\zeta)o}-1. \tag{19}$$

As long as operating and liquidity substitution costs in the two-tier system exceed pass-through savings, $\alpha^{\star \mathrm{p}} > \alpha^\star$ and $\alpha^{\mathrm{p}} > \alpha$.

Since we have no credible information about o we assume conservatively that pass-through costs are high, namely ten times the unit operating costs of reserves, $o = 10\rho$. Below, we consider alternative, even higher values. Figure 2 illustrates the results. Subject to optimal reserve holdings, CBDC operations would have to be between 0.24 and 14% more efficient than narrow banking to rationalize the introduction of CBDC when $\eta \to \infty$. Subject to actual reserve holdings, the required efficiency advantage lies between -44 and $+33\%$; more than 77% of α^{p} realizations are smaller than zero. The case for CBDC therefore is weak if reserve holdings are optimal but stronger if they correspond to actual ones.

When η is finite, the optimal composition of real balances is governed by a variant of equation (SP-4) that accounts for pass-through costs. With $\mu = \nu + \rho$ and $\lambda = 1$, the unit liquidity costs of CBDC and deposits equal $\nu + \rho + (1-\zeta)o$ and $\nu + \omega(\zeta^\star, \zeta^\star) + \zeta^\star\rho$, respectively. For $\eta = 10$ the optimal ratio $m^\star/n^\star$ then lies between 32 and 98%, and for $\eta = 50$ it lies between 0.38 and 89%. Typical values in the two scenarios are $3/4$ and $1/3$, respectively. That is, even with high pass-through costs, a finite elasticity of substitution implies a substantial optimal share of CBDC in real balances.

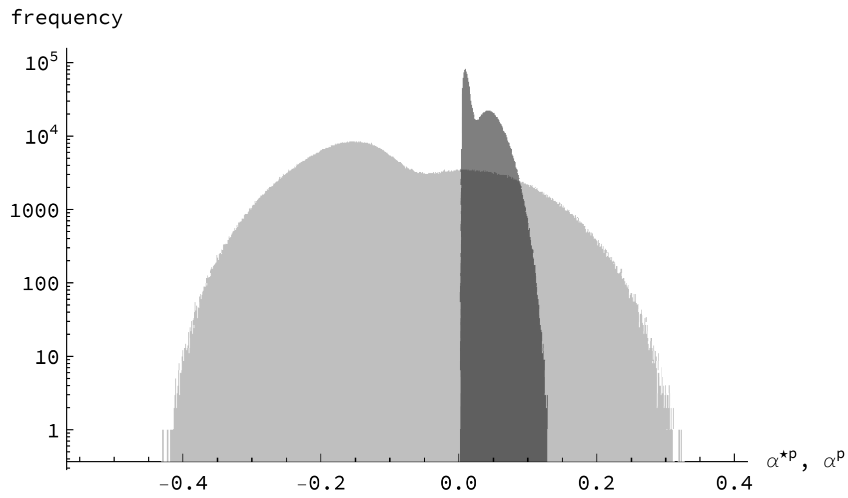

Figure 2: Histograms of $\alpha^{\star \mathrm{p}}$ (dark gray) and α^{p} (light gray) when $o = 10\rho$; $2'000'000$ draws each, $1'000$ bins, logarithmic scale.

4.5 Too-Big-To-Fail Banks and Tax Distortions

We conclude this section with two additional sets of frictions. Unlike pass-through costs, they raise the cost of liquidity provision by banks.

Too-Big-To-Fail Banks The first set of frictions concerns unobservable bank choices and lack of commitment on the part of government, which 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 private information. While both types have identical return characteristics, the newly introduced type generates private benefits for bank management but requires additional liquidity substitution, which costs $\hat{\gamma} > 0$ per deposit. This additional liquidity substitution is not contractible 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 $\omega_t(\zeta_{t+1}, \bar{\zeta}_{t+1}) + \hat{\gamma}$. Letting banks fail after insufficient liquidity substitution would have large social costs.

In equilibrium, banks choose to invest in the capital with private benefits but do not pay the additional costs $\hat{\gamma}$. This forces the government to bail banks out at social costs $\gamma > 0$ per deposit and raises the unit social costs in the two-tier system correspondingly. If deposits circulate in the first best (in which no γ costs arise), the first best ceases to be implementable. In any case, the two-tier system becomes less attractive:

Proposition 4. Information frictions, lack of commitment and too-big-to-fail banks change the government's cost-benefit calculus in favor of circulating CBDC.

Proof. The optimal policy subject to lack of commitment satisfies conditions (RA-1)–(RA-4) with $\nu + \gamma$ replacing ν. $\square$

While the cost and information structures underlying proposition 4 are quite specific the broader message of the proposition is robust and parallels the message of subsection 4.4: The case for CBDC must be assessed based on the actual costs of public and private liquidity provision, not those under ideal conditions. Information frictions, lack of commitment and too-big-to-fail banks push the actual costs of private provision above the ideal ones.

Tax Distortions The second friction concerns tax distortions or more generally, distortions that arise in the process of addressing the frictions in the banking sector, lack of competition and reserve externalities. We formalize this friction by assuming that taxing households (τ_t) causes deadweight burdens; ceteris paribus, lower taxes therefore increase welfare. To compare tax revenues and implied tax distortions in single- and two-tier systems, we contrast government revenues from sources other than taxes, namely 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

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

Under the Ramsey policy, two factors undermine budget balance: The reserves subsidy, $\chi^{r\star}_{t+1} - \rho = \omega_{\bar{\zeta},t}(\zeta^\star_{t+1}, \zeta^\star_{t+1}) < 0$, and the deposit subsidy, $\theta^\star_t$. Summing the two yields, from conditions (SP-1), (RA-3) and (RA-4),

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

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 deposit subsidy; the net budgetary impact then reflects the other component of the deposit subsidy, which counteracts monopsonistic price setting.

In conclusion, a single-tier payment system requires lower taxes and thus generates fewer tax distortions than a two-tier system—even outside crisis periods. If the government used nonfiscal instruments to address market failure in the banking sector, the two-tier system would still generate larger distortions, for example due to evasion and enforcement efforts. In either case, therefore, the need for corrective intervention in the two-tier system changes the social cost-benefit calculus in favor of a single-tier system.

This does not imply that CBDC should be issued; it should not, if the resource costs of a deposit based system are sufficiently low and money and deposits are perfect substitutes. But even in this case, the distortions caused by corrective interventions can rationalize a noncirculating CBDC that disciplines banks. Recall that targeting a sufficiently high interest rate on CBDC (without actually issuing it in equilibrium) can replicate the incentive effects of a deposit subsidy (proposition 3). 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 may offer welfare gains.

Proposition 5. Distortions caused by corrective interventions in the two-tier system change the Ramsey government's cost-benefit calculus in favor of CBDC. Distortions and a sufficiently high $\theta_t^\star$ can rationalize a noncirculating CBDC in place of deposit subsidies.

Quantitative Implications To account for tax distortions in addition to pass-through costs, we modify the $\alpha^{\star\mathrm{p}}$ statistic in (18) to

$$\alpha^{\star\mathrm{pt}} \equiv \frac{\nu+\rho}{\nu+\omega(\zeta^\star,\zeta^\star)+\zeta^\star\rho-(1-\zeta^\star)o+X}-1,$$

where X denotes the excess burden due to tax distortions per deposit. A parallel correction in the denominator of the α^{p} statistic in (19) yields the statistic α^{pt}, which is based on the actual rather than optimal reserves-to-deposits ratio. Under assumption 1, the Ramsey analysis implied fiscal costs per deposit of

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

when the composition of real balances is targeted. The denominator in the expression for $\alpha^{\star\mathrm{pt}}$ can then be expressed as

$$(\nu+\omega(\zeta^\star,\zeta^\star)+\zeta^\star\rho)\,(1+\psi x)-(1-\zeta^\star)o,$$

where x represents the excess burden of a tax dollar. In the expression for α^{pt}, we make a parallel substitution. As long as operating and liquidity substitution costs in the two-tier system exceed pass-through savings, $\alpha^{\star\mathrm{pt}} < \alpha^{\star\mathrm{p}}$ and $\alpha^{\mathrm{pt}} < \alpha^{\mathrm{p}}$.

As before, we let $\psi = 1/3$ and $o = 10\rho$ (below we consider alternative, higher o values) and we assume that $x = 0.25$ (Saez et al., 2012).[43] Figure 3 illustrates the resulting distributions of $\alpha^{\star\mathrm{pt}}$ and α^{pt}. Compared with figure 2, which illustrated the case with pass-through costs but no distortions, both distributions are shifted to the left. Statistic $\alpha^{\star\mathrm{pt}}$ lies between -7.5 and $+3.5\%$ with mean and median below -5%; the cumulative density evaluated at $\alpha^{\star\mathrm{pt}} = 0$ exceeds 99%. Statistic α^{pt} lies between -48 and $+18\%$ with typical values around -19%; its cumulative density evaluated at $\alpha^{\mathrm{pt}} = 0$ exceeds 94%. These results suggest that CBDC issuance subject to the operating cost structure of narrow banks (or cheaper) generates social cost savings relative to deposits, independent of whether reserves are at optimal or actual levels.

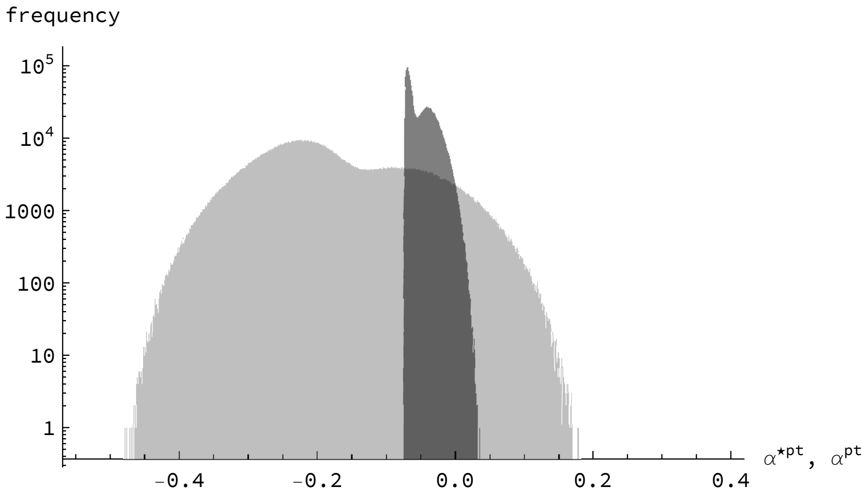

Figure 3: Histograms of $\alpha^{\star\mathrm{pt}}$ (dark gray) and α^{pt} (light gray) when $o = 10\rho$ and $x = 0.25$; $2'000'000$ draws each, $1'000$ bins, logarithmic scale.

When CBDC and deposits are imperfect substitutes ($\eta < \infty$), the optimal composition of real balances reflects the substitution elasticity of households in addition to the unit costs of the two sources of liquidity. With $\mu = \nu + \rho$ and $\lambda = 1$, the unit costs equal $\nu+\rho+(1-\zeta)o$ for CBDC and $(\nu+\omega(\zeta^\star, \zeta^\star)+\zeta^\star\rho)(1+\psi x)$ for deposits, and the optimal ratio $m^\star/n^\star$ always lies above 73% when $\eta = 10$ and above 21% when $\eta = 50$, with means and medians around 1.7 in the former case and around 18 in the latter. That is, optimal policy typically calls for substantially more CBDC than deposits even with a finite elasticity of substitution. Intuitively, the technological advantage of fractional reserve banking over

[43] For simplicity, we disregard effects of tax distortions on $\zeta^\star$.

narrow banking and the social costs of pass-through funding are minor compared to the excess burden of measures to address frictions in the banking sector.

These results hold under the—conservative—assumption of $o/\rho = 10$ and, as shown above, depend on the stipulated elasticity η. Figure 4 allows to assess the role of these two parameters more globally. For each $(o/\rho, \eta)$ pair, we consider the median of the implied $m^\star/n^\star$ realizations, and in figure 4 we plot the median's contour levels in $(o/\rho, \eta)$ space. The figure shows that a key threshold value for o/ρ is roughly 38: Independent of η, the optimal composition of real balances is tilted in favor of CBDC ($m^\star/n^\star \geq z$ for some $z \geq 1$ in 50% of draws) until the unit pass-through costs o exceed roughly 38ρ, which we consider very high. How quickly deposits optimally crowd out CBDC when unit pass-through costs increase further depends on the elasticity of substitution. If the elasticity is very high, the median of the $m^\star/n^\star$ realizations falls rapidly. If the elasticity is lower, the response is more measured; for example, with $\eta = 25$ and $o/\rho = 70$ the optimal composition $m^\star/n^\star$ still exceeds 20% in half of the realizations.

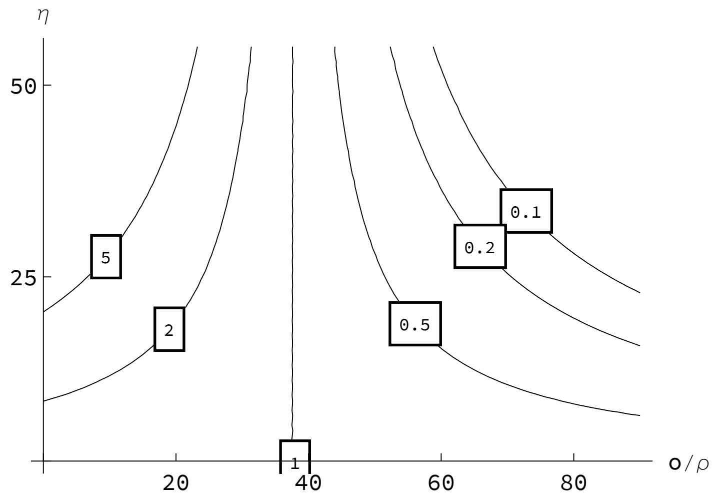

Figure 4: Contour plot of the median of $m^\star/n^\star$ for different $(o/\rho, \eta)$ combinations when $\mu = \nu + \rho$; 2′000′000 draws. Numbers in boxes indicate contour levels.

We draw two main conclusions from the analysis of this most comprehensive setting. First, as found before, it is critical to account for indirect in addition to direct social costs and benefits when ranking monetary architectures. And second, the various costs and benefits considered here point to an important role of CBDC in an optimal monetary architecture. Our calibrations suggest a dominant share of CBDC in real balances unless pass-through costs are very high. Even in this case, the optimal CBDC share remains substantial unless CBDC and deposits are very close substitutes.

5 Conclusion

We have analyzed optimal monetary architecture—single-tier, two-tier or mixed—and optimal monetary policy. Regarding the latter, our findings blend a generalized version of the Friedman (1969) rule with Pigou (1920): The terms at which liquidity is provided should reflect social costs accounting for externalities. Consequently, CBDC and reserves should be remunerated differently. When certain policy instruments are not admissible or their deployment causes distortions, there can also be a case for noncirculating CBDC as a substitute.

The optimal monetary architecture reflects social costs and household preferences. From a strict payment operations perspective, a two-tier architecture dominates a single-tier system with a narrow-bank like cost structure. Essential and costly pass-through funding from the central bank to banks strengthens this superiority. But lack of competition in deposit markets, externalities from banks' reserve holdings and social costs of addressing these and other frictions in the two-tier system erode it. All things considered, our calibrated model robustly suggests a cost advantage of CBDC unless pass-through funding is essential and very costly. How this advantage translates into optimal portfolio shares depends on the degree of substitutability between CBDC and deposits.[44]

Factors we have abstracted from could alter these conclusions. One potentially important channel the model does not capture relates to innovation. If banks that deliver CBDC payment services on behalf of the central bank lack incentives to innovate, then this could weaken the case for CBDC. Another factor are fixed costs of CBDC deployment. It appears unlikely that such costs could overturn our quantitative results based on variable costs except if CBDC infrastructure depreciated quickly. But fixed costs are potentially more relevant when CBDC only serves as a disciplining device rather than an actually circulating instrument.

Yet other factors concern pass-through costs. We view pass-through costs that imply a CBDC cost advantage as plausible. But a change of monetary architecture leads into uncharted territory where longer central bank balance sheets and pass-through funding might generate costs whose size we underestimate. For example, a CBDC based architecture could require costlier defenses against cyber attacks.[45] Or, as discussed in

[44]Technological progress has reduced the costs of payment operations (Hanson et al., 2015; Danmarks Nationalbank, 2018b; Junius et al., 2022) and will likely continue to do so. This does not alter our conclusions as long as the cost components entering the $\alpha^{\star\mathrm{pt}}$ and α^{pt} statistics are symmetrically affected.

[45]Eisenbach et al. (2022) argue that an attack on a major bank currently risks impairing a third of the payment network. A more centralized CBDC based architecture could change that share and the cost of preventive action. Doerr et al. (2022) assess financial sector cyber risks.

subsection 4.4, information frictions could aggravate the consequences of central bank mistakes.[46] Further research is required to better understand the role of these and other factors.

Our workhorse model can serve as the basis for many useful extensions. The positive analysis deserves further exploration of the transmission mechanism in closed and open economies. The representative agent assumption could be relaxed to analyze distributive conflict, e.g., across generations or between "bankers" and "nonbankers", and its resolution in politico-economic equilibrium. Fiscal and monetary policy could be separated to study strategic interaction between different government players. And the flexible price assumption could be replaced by stipulated nominal rigidities to analyze monetary policy subject to broader objectives and a richer constraint set.

[46] However, Keister and Monnet (2022) emphasize that CBDC enlarges central bank information sets.

A Cash and Government Bonds

In this appendix we argue that abstracting from cash and government bonds does not undermine the generality of the analysis.

Consider cash first. It would enter the household's program as another argument of the z function; its interest rate would equal the deflation rate; and like money and deposits, it would generate resource costs when used as a means of payment. The new insights gained would mainly concern cash-CBDC substitution, a swap of central bank liabilities without major macroeconomic consequences.[47] Our framework instead focuses on the "disruptive" deposit-CBDC substitution of relevance for the banking sector and the macro economy. If prices were sticky, cash could be macro economically relevant by giving rise to a binding effective lower bound on interest rates.[48]

Including government bonds in the analysis would also have only minor effects. Government bonds as central bank assets would be irrelevant because only the consolidated government budget constraint matters. Moreover, with nondistorting taxes and homogeneous households, government debt would leave the shadow value of public funds unchanged. Government bonds would also be irrelevant in household balance sheets, except possibly as hedging instrument if returns on money, deposits and bonds varied differentially across contingencies, which is of limited relevance empirically.

But if bonds do not offer hedging benefits for households then banks do not have an incentive to hold bonds either.[49] Bonds might be useful as bank assets, however, if they provided liquidity services, for instance as collateral in repo transactions. In the model, reserves already play that role.[50] 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.

B Costs of Liquidity Substitution

In this appendix we micro found function ω_t and derive the portfolio choice of banks under assumption 2.

[47] Keister and Sanches (2023) find minor effects of a "cash-like" CBDC.

[48] Cash may also affect the market power of banks (Drechsler et al., 2017; Lagos and Zhang, 2022) and thus the elasticity of deposit funding. This mechanism operates independently of CBDC and the model does not restrict the elasticity.

[49] Households could replicate the effect of a bank's bond purchases by holding bonds themselves.

[50] Empirically, banks hold few treasury securities but they do hold agency mortgage-backed securities (Hanson et al., 2015).

B.1 Micro Foundations

We propose two micro foundations. The first is based on a fire sale narrative; it yields the specification imposed under assumption 2, which in turn constitutes a special case of the function ω_t imposed throughout the analysis. The second relates to frictional interbank markets.

Fire Sales 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, 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 imply a cap on 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 depositors, 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.[51] 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 outset as a second type of bank asset. Second, we emphasize bank liquidity rather than solvency and interpret fire sales of capital as operations to reduce a liquidity squeeze. 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

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

indicating that deposit creation is constrained by worst case asset values. We replace this

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

inequality by the condition that deposits must be covered by the sum of reserves, fire-sale receipts $q_t k_{t+1}$ and the liquidity substitution afforded by the costly technology:

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

The inequality holds with equality if the bank minimizes costs; it can be written as

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

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

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

We assume that ϕ_t is sufficiently large for active banks to choose $\zeta_{t+1} > 0$. Since $\varphi > 1$ and $\bar{\varphi} > 0$, ω_t satisfies the conditions stipulated in the text; specifically, $\omega_{\zeta\zeta,t} + \omega_{\zeta\bar{\zeta},t} > 0$ and $\omega_{\zeta\zeta,t} + \omega_{\zeta\bar{\zeta},t} + \omega_{\bar{\zeta}\zeta,t} + \omega_{\bar{\zeta}\bar{\zeta},t} > 0$ in equilibrium.

Frictional Interbank Markets The functional form derived above has two key properties. First, a more liquid asset side of the balance sheet entails private benefits for a bank; this is captured by the assumption $\omega_{\zeta,t} < 0$ ($\varphi > 0$). Second, more liquid bank assets generate external benefits; this is captured by the assumption $\omega_{\bar{\zeta},t} < 0$ ($\bar{\varphi} > 0$). The conventional assumption of decreasing marginal benefits (both private and social) along the equilibrium path implies the conditions on second derivatives ($\varphi > 1$).

Costs of bank balance sheet management with these two properties are present in many other contexts. Frictional interbank markets are a prime example. Bianchi and Bigio (2022) analyze the problem of a bank that issues deposits and holds reserves as well as less liquid assets. A bank is subject to stochastic reserve in- and outflows; the end-of-period stock of reserves must not fall below a threshold. The bank has access to a frictional OTC interbank market for reserves, may deposit reserves at the central bank at a low rate or borrow them at the central bank's discount window at a high rate and subject to "stigma" costs.

Bianchi and Bigio (2022) derive a bank's "liquidity yield function," which characterizes the costs of replenishing reserves. They are the product of the reserve shortfall and a strictly positive cost term, which is strictly increasing in interbank market tightness (aggregate reserve demand by banks that need to borrow reserves relative to aggregate supply by banks willing to lend). This is in line with our assumption that $\omega_t n_{t+1}$ is strictly positive unless $\zeta_{t+1} = 1$ and that ω_t is decreasing in the bank's own reserve holdings as well as in aggregate reserve holdings. It is also consistent with the specific functional form derived above, which may reflect a probability of having to replenish reserves that increases in the bank's share of illiquid asset holdings as well as costs of replenishing that are decreasing in aggregate reserves. Either formulation captures the fact that precautionary reserve holdings reduce a bank's need to engage in costly measures to make up for a shortfall.[52] As a consequence, both Bianchi and Bigio (2022) and we find that monetary policy affects the supply of bank credit.

Evidence The cost function ω_t in general and the specific functional form derived above in particular imply a negatively sloped equilibrium demand curve for reserves, $\chi^r_{t+1}(\zeta_{t+1})$, see below. This is consistent with empirical evidence on the closely related demand function in the interbank market, $\bar{\chi}^r_{t+1}(\bar{\zeta}_{t+1})$ say. Afonso, Giannone, La Spada and Williams (2022) estimate derivatives of the latter function and fit a function that is concave for low values of $\bar{\zeta}_{t+1}$ and convex for high ones. The shape of our χ^r_{t+1} function conforms with the more relevant convex part of the fitted $\bar{\chi}^r_{t+1}$ function.[53]

[52] Bianchi and Bigio (2022) explicitly model bank heterogeneity; their functional form assumptions are guided by the motivation to admit aggregation.

[53] For additional evidence see, for example, Bianchi and Bigio (2022) or Afonso, Duffie, Rigon and Shin (2022).

B.2 Implied Bank Choices Under Assumption 2

The bank's optimality condition for reserves, equation (8), implies

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

In equilibrium,

$$\zeta_{t+1} = \bar{\zeta}_{t+1} = 1 - \left(\frac{\chi^r_{t+1}}{\phi_t \varphi}\right)^{\frac{1}{\varphi+\bar{\varphi}-1}}. \tag{8'}$$

We assume that ϕ_t is sufficiently large for the equilibrium choice of ζ_{t+1} to be strictly positive; such a large value for ϕ_t always exists. Accordingly, the equilibrium costs of liquidity substitution equal

$$\omega_t(\zeta_{t+1}, \zeta_{t+1}) = \phi_t \left(\frac{\chi^r_{t+1}}{\phi_t \varphi}\right)^{\frac{\varphi+\bar{\varphi}}{\varphi+\bar{\varphi}-1}} \tag{20}$$

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 (9)),

$$\tilde{\omega}_t(-\chi^r_{t+1}) = \chi^r_{t+1} - (\varphi - 1)\phi_t \left(\frac{\chi^r_{t+1}}{\phi_t \varphi}\right)^{\frac{\varphi+\bar{\varphi}}{\varphi+\bar{\varphi}-1}}. \tag{21}$$

C Equilibrium Deposit Spread

Suppose that assumption 1 holds and 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

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

When deposits circulate, the Euler equation (3) reduces to

$$\frac{\vartheta z_{t+1}^{-\psi}}{(1-\vartheta)c_t^{-\psi}} \left(\frac{n_{t+1}}{z_{t+1}(1-\tilde{\lambda}_t)}\right)^{-\frac{1}{\eta}} = \chi^n_{t+1} = 1 - \frac{R^n_{t+1}}{R^f_{t+1}}.$$

If, moreover, the central bank targets $\xi_{t+1} = m_{t+1}/n_{t+1}$ (this includes the no-CBDC case where the central bank sets $m_{t+1} = 0$), then z_{t+1} is proportional to n_{t+1} and the condition simplifies to

$$\text{constant} \cdot \frac{n_{t+1}^{-\psi}}{c_t^{-\psi}} = 1 - \frac{R_{t+1}^n}{R_{t+1}^f}.$$

The elasticity of deposit funding with respect to the spread thus equals $-\psi^{-1}$ and the elasticity with respect to the deposit rate satisfies $\eta_{t+1}^n = (\psi \chi_{t+1}^n R_{t+1}^f / R_{t+1}^n)^{-1}$. Accordingly, equation (9) reads

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

or

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

D Calibration

We base the calibration on macro finance data and model implied equilibrium conditions. Unless noted otherwise, we use monthly data over the period January 2010 to December 2019. We report spreads and unit costs on an annual basis.

We use multiple data sources and allow for measurement error. As a consequence we obtain multiple calibrations. Rather than averaging or aggregating in other ways, we acknowledge the uncertainty about model parameters and assess the robustness of our findings based on it. When different data sources or potential measurement error imply a range of possible values for an input in the calibration, we treat that input as uniformly distributed between the minimum and maximum values. Drawing repeatedly from the distributions of the inputs, we derive a calibration for each input combination, compute the statistics of interest for each calibration and thus arrive at distributions of the statistics of interest.

Measuring χ^r, r, n We construct a risk-free interest rate series by adding a constant convenience yield of 65 bp to the FRED series `TB3MS` for the 3-month treasury bill rate. van Binsbergen et al. (2022) infer convenience yields from risky asset prices over the period 2004–2018; except for financial crisis periods they find a stable convenience yield for treasury bills of around 65 bp. We use the FRED series `IOER` for the interest rate on reserves. We use the two series and the definition $\chi_{t+1}^r \equiv 1 - R_{t+1}^r / R_{t+1}^f$ to construct a series for the liquidity premium on reserves. We obtain an average value $\chi^r = 0.4970 \cdot 10^{-2}$.

We use the FRED series `TOTRESNS` for reserves and the FRED series `DPSACBM027NBOG` for deposits. We use the two series and the definition $\zeta_{t+1} \equiv r_{t+1}/n_{t+1}$ to construct a series for the reserves-to-deposits ratio. We obtain an average value $\zeta = 0.1945$. We condition the calibration on the average values χ^r and ζ but allow for measurement error.

Calibrating ν Hanson et al. (2015) use a hedonic regression approach to identify the noninterest expenses of banks associated with deposit-taking. They find an (annual) cost ratio of 1.3%, which corresponds to ν in the model. Van den Heuvel (2022, p. 35) finds a marginal net noninterest cost ratio of 1.22%; combined with an estimated noninterest income ratio of 49 bp (Hanson et al., 2015), this implies a gross cost ratio of 1.71%. Wang et al. (2020, p. 25) find that "banks incur a 0.9% cost of maintaining deposits".

More generally, ν represents costs borne throughout the private sector, not only by banks. A comprehensive Norwegian study estimates social costs of domestic payment services in 2020 of 71 bp of mainland GDP when cash handling costs (8 bp) are excluded (Norges Bank, 2022). Normalizing by the ratio of transaction accounts to mainland GDP yields an annual cost ratio of between 0.75 and 1.79%, depending on whether transaction account balances of all sectors or households are considered.[54] Other studies over the last fifteen years yield estimates of the GDP share of payment costs of similar magnitude.[55]

Several central bank studies offer information about social costs as a share of the value of total payments. An aggregate estimate for Denmark yields 1%, with substantial variation across payment instruments (Danmarks Nationalbank, 2018b). A Finnish study finds cost ratios up to 56 bp (Sintonen and Takala, 2022). With a velocity of 1.3, this implies ν values between 0.73 and 1.3%.[56]

Not all these costs are variable costs. Depending on payment instrument and payment habits, the fixed-cost shares of payments vary between one and two thirds (Junius et al., 2022, p. 17). We use the smaller correction factor because only few "fixed-cost" components are truly fixed on a macroeconomic scale, in particular when changes in deposit balances are associated with variation in the number of payments rather than their

[54] Statistics Norway, tables `www.ssb.no/en/statbank/table/09189` and `www.ssb.no/en/statbank/table/11003`.

[55] See for example Danmarks Nationalbank (2018a), Junius et al. (2022, Table 9), Norges Bank (2022, Table 4) and Sintonen and Takala (2022, Table 7). Excluding cash handling costs and costs borne by households, Euro area social costs in the late 2000s amounted to 47 bp of GDP (Schmiedel et al., 2012, Table 6). In Australia in 2013, payments from consumers to merchants and financial institutions generated social costs of 54 bp of GDP (Stewart et al., 2014). The non-U.S. estimates likely provide lower bounds for the U.S. where the modernization of the payment system has progressed at a slower pace (Duffie, 2021).

[56] The velocity corresponds to the Euro area M1-velocity in 2018 (Benati, 2020).

average size.[57] Accordingly, we arrive at the bounds $0.73\%\frac{2}{3}$ and $1.79\%\frac{2}{3}$ or

$$\nu \sim U[0.4867, 1.1933] \cdot 10^{-2}.$$

Calibrating ω We impose assumption 2, $\omega(\zeta, \bar{\zeta}) = \phi(1-\zeta)^{\varphi}(1-\bar{\zeta})^{\bar{\varphi}}$, and use the observed ζ and χ^r as well as additional data to identify ω.

We derive a first moment from a measure of aggregate bank vulnerability (AV) constructed by Duarte and Eisenbach (2021). They find that without the post-crisis shift towards more liquid bank assets that occurred between the end of 2006 and late 2008, AV would have been 15% higher starting in late 2008 (see also Roberts et al., 2023). Since $1-\zeta$ fell by 7.44% during these two years, this implies an elasticity of AV with respect to $1-\zeta$ of 2.0159. We assume that AV is proportional to the liquidity substitution that banks would have had to engage in to make up for asset illiquidity, $\omega^{1/\varphi}n = \phi^{1/\varphi}(1-\zeta)(1-\bar{\zeta})^{\bar{\varphi}/\varphi}n$. In equilibrium, the elasticity of this measure with respect to $1-\zeta$ equals $1+\bar{\varphi}/\varphi$ and we therefore set

$$1 + \frac{\bar{\varphi}}{\varphi} = 2.0159 \cdot U[0.8, 1.2],$$

where we allow for measurement error of $\pm 20\%$. Figure 2 in Duarte and Eisenbach (2021) suggests elasticities of similar magnitude when we directly compare the evolution of AV and $1-\zeta$ during several time periods.

We derive a second moment from the measured liquidity premium on reserves. Recall the bank's first-order condition for reserves in equilibrium, $\phi\varphi(1-\zeta)^{\varphi+\bar{\varphi}-1} = \chi^r$, which implies an elasticity of χ^r with respect to $1-\zeta$ of $\varphi+\bar{\varphi}-1$. During the sample period this elasticity equals 1.8361 and we therefore set

$$\varphi + \bar{\varphi} - 1 = 1.8361 \cdot U[0.8, 1.2],$$

where we allow for measurement error of $\pm 20\%$.

For validation purposes, we also compute the elasticity of the net interest rate on reserves, $R^r_{t+1} - 1$, with respect to reserves. During the sample period, this elasticity averages -1.8439. This compares with results in Afonso, Giannone, La Spada and Williams (2022, Table 2) who structurally estimate a time-varying demand curve on the interbank market over the period 2010–2021 and obtain elasticities of the federal funds rate with

[57] Compliance costs constitute important cost components; in the U.S. financial sector, 90% of compliance costs are labor costs (Trebbi and Zhang, 2022). Assuming a lower variable-cost share for commercial and central banks would increase the relative importance of liquidity substitution costs and strengthen the case in favor of a single-tier system.

respect to reserves in the $(-2,0)$ range.

Finally, we use the bank's first-order condition to identify the intercept of the ω function. Evaluating the condition at the observed values ζ and χ^r subject to measurement error of $\pm 20\%$ yields

$$\phi\varphi(1-\zeta)^{\varphi+\bar{\varphi}-1} = \chi^r \quad \text{with} \quad \zeta = 0.1945 \cdot U[0.8, 1.2], \;\; \chi^r = 0.4970 \cdot 10^{-2} \cdot U[0.8, 1.2].$$

For each set of draws, the three preceding conditions imply a calibration for $(\phi, \varphi, \bar{\varphi})$.

Figure 5 displays the $\omega(\zeta, \zeta)$ functions implied by five hundred draws and figure 6 displays the associated demand functions $\chi^r(\zeta)$. (The vertical line in each figure indicates $\zeta = 0.1945$.) The average value of ϕ across $1'000'000$ calibrations slightly exceeds 50 bp and the average values of φ and $\bar{\varphi}$ are around 1.4; we have $\phi \in [0.2968, 0.8547] \cdot 10^{-2}$, $\varphi \in [1.0208, 1.9860]$, and $\bar{\varphi} \in [0.9385, 1.8789]$. These values satisfy the conditions imposed on function ω_t and they imply positive and decreasing benefits of reserves as well as positive externalities.[58]

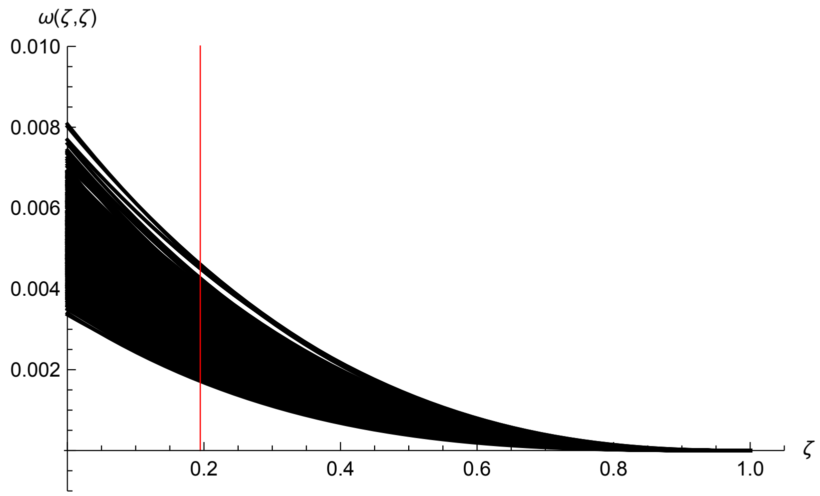

Figure 5: $\omega(\zeta, \zeta)$, 500 calibration draws.

The implied demand functions in figure 6 suggest that a liquidity premium on reserves in excess of 1% crowds out demand for reserves.[59] This is consistent with the fact that before the financial crisis interest on reserves equalled zero and so did, nearly, the reserves-to-deposits ratio. It is also consistent with much higher ex-post liquidity premia during

[58]The literature finds similarly shaped benefits of bank capital (e.g., Firestone et al., 2019); for the relation between bank solvency and liquidity, see for example Van den Heuvel (2022) and Roberts et al. (2023).

[59]The model implied demand function is convex (see appendix B), in line with the estimated relationship between the federal funds rate and reserves as long as reserves are not scarce (Afonso, Giannone, La Spada and Williams, 2022).

times of financial stress (Bianchi and Bigio, 2022; van Binsbergen et al., 2022). Bianchi and Bigio (2022) calibrate stigma costs of 5% and the literature finds even larger costs of fire sales, for instance after bankruptcy.[60] The calibrated ω can be interpreted as reflecting such costs during distress periods, weighted by their probability.

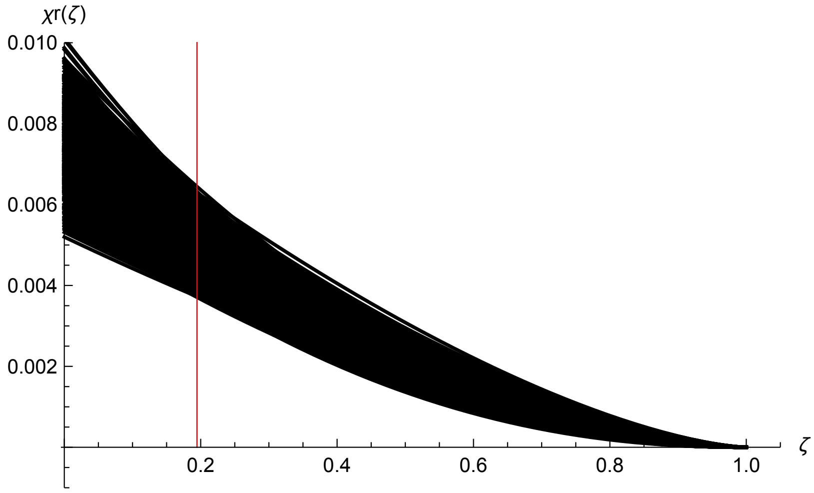

Figure 6: $\chi^r(\zeta)$, 500 calibration draws.

Calibrating ρ In 2021 the Federal Reserve System collected USD 456 million "revenues from priced services," corresponding to the fees charged for Fedwire transfer services; this amounts to less than 1.2 bp of reserves. By law these fees must cover costs.[61] Among U.S. large-value payment systems Fedwire accounts for 61% of processed transactions in terms of volume and 69% in terms of value (2021 data).[62] Scaling the Fedwire costs accordingly and accounting for fixed costs (1/3, as with ν) we arrive at

$$\rho \sim 0.1204 \cdot 10^{-3} \cdot U[0.8, 1.2],$$

where we allow for $\pm 20\%$ measurement error.[63]

[60] Bennet and Unal (2015) report discounted losses on the disposition of assets of failed banks of 23% of assets and total resolution costs of 33%. Based on FDIC data, Corbae and D'Erasmo (2021) stipulate loan portfolio liquidation costs of 20%. Based on debt recovery rates and FDIC resolution costs, Begenau and Landvoigt (2022) calibrate the share of assets lost in bankruptcy to 20 or 35%.

[61] See `www.federalreserve.gov/newsevents/pressreleases/files/other20221103a1.pdf`. Total system operating costs exceed the revenue from priced services by an order of magnitude.

[62] See `stats.bis.org/statx/srs/table/T8?c=US`, `stats.bis.org/statx/srs/table/T9?c=US`. The BIS "red book" surveys the U.S. payment system, see `www.bis.org/cpmi/publ/d105_us.pdf`. CHIPS, the main private large-value payment system, is run by The Clearing House, which is bank owned.

[63] Reserves might be created in the context of a sequence of actions by separate government entities. For example, Treasury might issue bonds, which the Fed subsequently purchases against reserves. This does not affect the net costs and benefits of reserves for the government or the private sector.

As with costs incurred by the private sector (represented by ν), financial statements may not fully reflect the social costs of reserve based payments. For a broader perspective, we turn again to central bank studies, which provide information about the share of payment operating costs borne by the central bank, $\rho\zeta/(\rho\zeta+\nu)$. Schmiedel et al. (2012) report that 3% of social costs related to retail payments are borne by the central bank, and Junius et al. (2022) summarize the findings of more recent studies, which find central bank cost shares in excess of 1.65%. However, substantial central bank costs are due to cash operations, which we abstract from. Sintonen and Takala (2022) report that almost all central bank costs in Finland are cash related while Norwegian data suggests a noncash central bank cost share of around 72 bp (Norges Bank, 2022). We rely on the latter estimate; allowing for ±20% measurement error, we let

$$\rho = \frac{\nu}{\zeta}\frac{\text{cost share}}{1-\text{cost share}}, \quad \text{cost share} \sim 0.7178 \cdot 10^{-2} \cdot U[0.8, 1.2].$$

Figure 7 illustrates the results of the two approaches. The Fedwire calibration implies ρ values at the lower end of the distribution implied by central bank comprehensive cost shares. For our analysis, we pool the results of the two approaches.

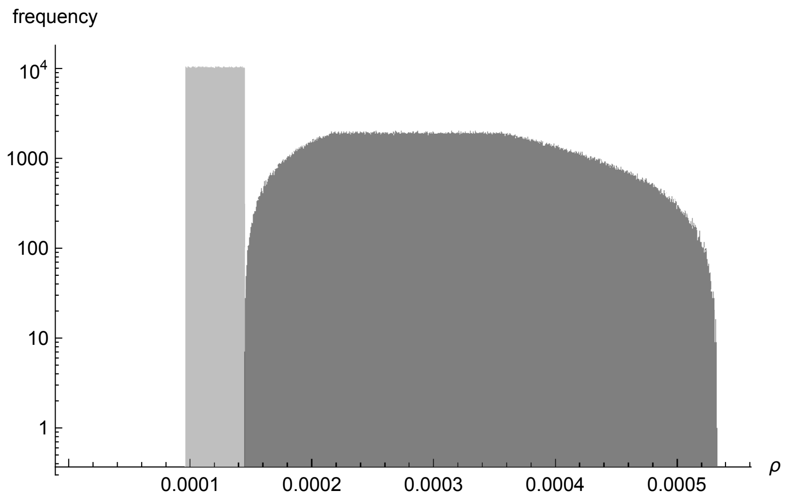

Figure 7: Histogram of ρ realizations; calibration based on Fedwire costs (light gray) and central bank cost shares (dark gray); 1′000′000 draws each, 1′000 bins, logarithmic scale.

We note that independent of the calibration approach, ρ is more than an order of magnitude smaller than ν. We also note that the typical ρ value in combination with the ω function calibrated above implies a large optimum reserves-to-deposits ratio. In fact, $\zeta^{\star}$ is much larger than its measured equilibrium counterpart, $\zeta = 0.1945$. The first approach to calibrating ρ implies values for $\zeta^{\star}$ in excess of 85% (0.9264 on average) and the second approach values in excess of 74% (0.8781 on average); the associated $\omega^\star$ values are on the order of 10^{-5}.

Importantly, this does not indicate problems with the calibration of ω. The large discrepancy between equilibrium and optimum ζ we find is independent of the specification and calibration of the ω function. Its cause is the difference between χ^r and ρ and the sign of that difference: According to the model, an optimum policy should set $\chi^r \leq \rho$ (see equations (SP-1) and (RA-3)) but the measured χ^r exceeds ρ by more than an order of magnitude. Even if we stipulated zero externalities of reserve holdings and assumed an ω function that best fits the data, the fact would remain that $\rho \ll \chi^r$ while theory predicts the reverse inequality under an optimal policy. While the model explains the observed ζ conditional on the observed χ^r, it does not explain the actual policy choice of χ^r.[64]

Table 1 in the main text summarizes the calibration.

E Optimality Under Assumptions 1 and 2

From condition (SP-1), the optimal reserves-to-deposits ratio, $\zeta^\star_{t+1}$, solves $\omega_{\zeta,t}(\zeta_{t+1}, \zeta_{t+1}) + \omega_{\bar{\zeta},t}(\zeta_{t+1}, \zeta_{t+1}) = -\rho$. Under assumption 2, this implies

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

such that

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

(Our assumption on ϕ_t guarantees that $\zeta^\star_{t+1} \in (0, 1)$.) The optimal liquidity premium on reserves thus equals

$$-\omega_{\zeta,t}(\zeta^\star_{t+1}, \zeta^\star_{t+1}) = \rho \frac{\varphi}{\varphi + \bar{\varphi}}.$$

From condition (RA-2), optimality requires

$$\chi^{n\star}_{t+1} = \nu + \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \zeta^\star_{t+1}\rho.$$

Under assumption 1 and when the central bank targets ξ_{t+1} (a special case of which is

[64] If we assumed that the measured χ^r corresponds to the Ramsey policy, the calibrated ω function would imply a value for ρ near 1%, which is inconsistent with the evidence.

setting $m_{t+1} = 0$), the equilibrium deposit spread satisfies (see appendix C)

$$\chi^n_{t+1} = \frac{\nu + \tilde{\omega}_t(-\chi^r_{t+1}) - \theta_t}{1 - \psi}.$$

Using the fact that $\tilde{\omega}_t(-\chi^{r\star}_{t+1}) = \omega_t(\zeta^\star_{t+1}, \zeta^\star_{t+1}) + \chi^{r\star}_{t+1}\zeta^\star_{t+1}$, this implies the optimal subsidy

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

References

Afonso, G., Duffie, D., Rigon, L. and Shin, H. S. (2022). How abundant are reserves? Evidence from the wholesale payment system, *Working Paper 30736*, NBER, Cambridge, Massachusetts.

Afonso, G., Giannone, D., La Spada, G. and Williams, J. C. (2022). Scarce, abundant, or ample? A time-varying model of the reserve demand curve, *Staff Report 1019*, Federal Reserve Bank of New York, New York.

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

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

Auer, R., Cornelli, G. and Frost, J. (2020). Rise of the central bank digital currencies: Drivers, approaches and technologies, *Working Paper 880*, Bank for International Settlements, Basel.

Barrdear, J. and Kumhof, M. (2022). The macroeconomics of central bank digital currencies, *Journal of Economic Dynamics and Control* **142**: 104148.

Begenau, J. and Landvoigt, T. (2022). Financial regulation in a quantitative model of the modern banking system, *Review of Economic Studies* **89**(4): 1748–1784.

Benati, L. (2020). Money velocity and the natural rate of interest, *Journal of Monetary Economics* **116**: 117–134.

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

Bennet, R. L. and Unal, H. (2015). Understanding the components of bank failure resolution costs, *Financial Markets, Institutions & Instruments* **24**(5): 349–389.

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. (2022). Banks, liquidity management, and monetary policy, *Econometrica* **90**(1): 391–454.

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.

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

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.

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, S. M., Jiang, J. and Zhu, Y. (2023). Bank market power and central bank digital currency: Theory and quantitative assessment, *Journal of Political Economy* **131**(5): 1213–1248.

Corbae, D. and D'Erasmo, P. (2021). Capital buffers in a quantitative model of banking industry dynamics, *Econometrica* **89**(6): 2975–3023.

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.

Danmarks Nationalbank (2018a). The aggregate costs of payments in Denmark were kr. 15.6 billion in 2016. Series: Costs of Payments in Denmark 2016.

Danmarks Nationalbank (2018b). The costs of consumer-to-business payments have decreased considerably. Series: Costs of Payments in Denmark 2016.

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.

Doerr, S., Gambacorta, L., Leach, T., Legros, B. and White, D. (2022). Cyber risk in central banking, *Working Paper 1039*, Bank for International Settlements, Basel.

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

Duarte, F. and Eisenbach, T. M. (2021). Fire-sale spillovers and systemic risk, *Journal of Finance* **76**(3): 1251–1294.

Duffie, D. (2021). Building a stronger financial system: Opportunities for a Digital Dollar, *in* D. Niepelt (ed.), *Central Bank Digital Currency: Considerations, Projects, Outlook*, CEPR Press, London, chapter 17, pp. 137–146.

Eisenbach, T. M., Kovner, A. and Junho Lee, M. (2022). Cyber risk and the U.S. financial system: A pre-mortem analysis, *Journal of Financial Economics* **145**(3): 802–826.

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

Firestone, S., Lorenc, A. and Ranish, B. (2019). An empirical economic assessment of the costs and benefits of bank capital in the United States, *Federal Reserve Bank of St. Louis Review* **101**(3): 203–230.

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

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.

Junius, K., Devigne, L., Honkkila, J., Jonker, N., Kajdi, L., Kimmerl, J., Korella, L., Matos, R., Menzl, N., Przenajkowska, K., Reijerink, J., Rocco, G. and Rusu, C. (2022). Costs of retail payments—an overview of recent national studies in Europe, *Occasional Paper 294*, European Central Bank, Frankfurt.

Keister, T. and Monnet, C. (2022). Central bank digital currency: Stability and information, *Journal of Economic Dynamics and Control* **142**: 104501.

Keister, T. and Sanches, D. (2023). Should central banks issue digital currency?, *Review of Economic Studies* **90**(1): 404–431.

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

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.

Lagos, R. and Zhang, S. (2022). The limits of ONETARY ECONOMICS: On money as a constraint on market power, *Econometrica* **90**(3): 1177–1204.

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. (2022). Money and banking with reserves and CBDC, *Discussion Paper 22-12*, University of Bern.

Norges Bank (2022). Costs in the Norwegian payment system 2020, *Norges Bank Papers 3*, Norges Bank, Oslo.

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.

Pasqualini, A. (2021). Markups, markdowns and bankruptcy in the banking industry. Unpublished, University of Oxford.

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.

Pigou, A. C. (1920). *The Economics of Welfare*, MacMillan and Co., London.

Roberts, D., Sarkar, A. and Shachar, O. (2023). Liquidity regulations, bank lending, and fire-sale risk, *Staff Report 852*, Federal Reserve Bank of New York, New York.

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

Saez, E., Slemrod, J. and Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review, *Journal of Economic Literature* **50**(1): 3–50.

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.

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

Sintonen, M. and Takala, K. (2022). Costs of retail payments in Finland: What paying costs?, *Expository Studies A129*, Bank of Finland, Helsinki.

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

Stewart, C., Chan, I., Ossolinski, C., Halperin, D. and Ryan, P. (2014). The evolution of payment costs in Australia, *Research Discussion Paper 2014-14*, Reserve Bank of Australia, Sydney.

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

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.

Trebbi, F. and Zhang, M. B. (2022). The cost of regulatory compliance in the United States, *Working Paper 30691*, NBER, Cambridge, Massachusetts.

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 Binsbergen, J. H., Diamond, W. F. and Grotteria, M. (2022). Risk-free interest rates, *Journal of Financial Economics* **143**(1): 1–29.

Van den Heuvel, S. J. (2022). The welfare effects of bank liquidity and capital requirements, *Finance and Economics Discussion Series 2022-72*, Federal Reserve Board, Washington, D.C.

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

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. (2022). Central bank digital currency: Welfare and policy implications, *Journal of Political Economy* **130**(11): 2829–2861.