Why Bank Money Creation?*

Hans Gersbach
Center of Economic Research
at ETH Zurich and CEPR
Zürichbergstrasse 18
8092 Zurich
Switzerland
hgersbach@ethz.ch

Sebastian Zelzner
Center of Economic Research
at ETH Zurich
Zürichbergstrasse 18
8092 Zurich
Switzerland
szelzner@ethz.ch

This version: June 15, 2022

Abstract

We provide a rationale for bank money creation in our current monetary system by investigating its merits over a system with banks as intermediaries of loanable funds. The latter system could result when CBDCs are introduced. In the loanable funds system, households limit banks' leverage ratios when providing deposits to make sure they have enough "skin in the game" to opt for loan monitoring. When there is unobservable heterogeneity among banks with regard to their (opportunity) costs from monitoring, aggregate lending to bank-dependent firms is inefficiently low. A monetary system with bank money creation alleviates this problem, as banks can initiate lending by creating bank deposits without relying on household funding. With a suitable regulatory leverage constraint, the gains from higher lending by banks with a high repayment pledgeability outweigh losses from banks which are less diligent in monitoring. Bank-risk assessments, combined with appropriate risk-sensitive capital requirements, can reduce or even eliminate such losses.

Keywords: monetary system, banking, money creation, loanable funds, capital requirements, leverage constraint, asymmetric information, moral hazard, CBDC

JEL Classification: E42, E44, E51, G21, G28

*We are grateful to Hugo van Buggenum, Hyun Song Shin, Pedro Pérez Velasco and seminar participants at ETH Zurich for helpful comments.

1 Introduction

The current monetary architecture has often attracted criticism, especially for its "magic money tree", which allows banks to create money "out of thin air": they can create claims on the legal tender banknotes in the form of deposits, which are the main source of money in our modern economies and are used by banks to grant loans or purchase assets from non-banks. Popular concerns that commercial banks then have access to an inexhaustible source of profits, as well as fears about financial stability have triggered so-called "sovereign money" initiatives to abolish this privilege of banks.[1] In parallel, central banks around the globe are considering the introduction of a central bank digital currency (CBDC). To what extent such a CBDC would impact commercial banks' current role in money creation is not clear yet. If a CBDC were to become the dominant medium of exchange and private bank deposits were to be moved into CBDC, banks could lose their money creation privilege and be reduced to simple intermediaries of loanable funds.

In this paper, we examine whether there is an economic rationale for our current two-tier monetary architecture with bank money creation, which essentially works as follows.[2] To a large extent, the money stock available to the public is composed of deposits (electronic private bank money) at commercial banks. Deposits are issued by commercial banks, in particular when they grant loans. Claims arising from interbank deposit flows—when the public makes payments—are settled by reserves (electronic central bank money) issued by the central bank (CB) to commercial banks. Importantly, banking regulation ensures that commercial banks comply with a set of rules such as capital requirements. We compare this two-tier monetary architecture with bank money creation (henceforth, MC economy) to the corresponding, standard loanable funds economy (henceforth, LF economy), in which banks need to acquire investment goods before they can grant loans to firms for capital investments.

Our main insights are as follows. In the LF economy, it is in the interest of households to limit banks' leverage ratios when providing deposits to ensure that banks have enough "skin in the game" to monitor their loans. When banks are het-

[1] In 2018, Switzerland voted on the "Vollgeld-Initiative", which aimed at doing that. See `https://www.vollgeld-initiative.ch/english/`. The proposal was rejected.

[2] For a more detailed analysis of the current monetary system, see Faure and Gersbach (2021).

erogeneous with regard to the (opportunity) costs of monitoring and when there is asymmetric information between households and banks about these characteristics, aggregate lending to bank-dependent firms is inefficiently low. In contrast, banks in the MC economy can initiate lending by creating bank deposits, without relying on household funding. With a suitable regulatory leverage constraint, the gains from higher lending by banks with a high repayment pledgeability outweigh losses from banks which are less diligent in monitoring. Bank-risk assessments, combined with appropriate risk-sensitive capital requirements, can reduce or even eliminate such losses, since these banks anticipate that high initial lending and leverage will not pass the regulatory requirements when the risk of their credit portfolio is assessed. If risk-assessment is perfect, the first-best allocation can be achieved in the MC economy.

At a more detailed level, we start with a two-period, two-sector economy with risk-neutral agents as in Gersbach and Rochet (2012, 2017), extended to heterogeneous banks and with asymmetric information of households about individual bank characteristics. Households and bankers are endowed with a capital good, which they supply to firms in two sectors in order to produce a consumption good. In the first sector of the economy, firms have direct access to the capital good through issuing bonds to households. In the second sector, firms can only obtain capital through bank loans. Banks partly finance their loans through their own endowment with capital goods, i.e., through equity, and partly either by household funding (in the LF economy) or by money creation (in the MC economy). Banks are subject to moral hazard in the spirit of Holmström and Tirole (1997). If they monitor loans diligently, their investments are more likely to succeed. If they shirk monitoring, they enjoy private benefits. Banks are heterogeneous regarding the benefits from shirking or, equivalently, regarding their efficiency in monitoring.

In the LF economy, the amount of funding households are willing to provide to banks is limited, since banks' monitoring incentives decrease proportionally to external financing and thus to the scale of the bank. With heterogeneous banks and asymmetric information between households and banks, households limit funds to banks, such that even the bank with the greatest potential benefits from shirking still monitors. As a consequence, aggregate external financing of banks, and thus aggregate lending by banks, is low. It is, of course, lower than in a first-best world without any frictions and it turns out that it is inefficiently low since in the MC economy and with the same informational frictions, aggregate bank lending will be higher, resulting in larger expected output overall.

In the MC economy, banks do not require household funding to initiate lending. Any loan they hand out simultaneously creates a deposit for the borrower. Firms use the deposits obtained through loans to buy the capital good from households, which are credited with deposits at their bank in return. As firms and households are likely to hold accounts at different banks, the ensuing interbank transactions have to be settled by reserves. Only banks can borrow such reserves from the CB.

As long as the profits on new loans exceed the bank's funding costs, increasing money creation, and thus leverage, is always profitable for an individual bank in the MC economy, since it increases the bank's expected return on equity. High leverage, however, implies low monitoring incentives. Hence, the government acting as a bank regulator imposes a leverage constraint.[3] By setting this leverage constraint, the regulator aims to strike an optimal balance between maintaining the banks' monitoring incentives on the one side and allowing an efficient allocation of capital on the other side. Put differently, the regulator faces a trade-off when deciding on the optimal leverage constraint: a tight constraint incentivizes monitoring, also at banks with a high exposure to moral-hazard, but leads to lower than optimal lending levels for diligent banks. If the regulator sets a sufficiently strict leverage constraint, all banks monitor and the resulting capital allocation is the same as in the LF economy. We show that selecting a somewhat looser leverage constraint improves economic outcomes. It implies that a positive fraction of banks shirks monitoring, but it also leads to a more efficient allocation of capital and, overall, to higher aggregate output than in the LF economy.

We also explore how the allocation in the MC economy can be further improved by risk-sensitive leverage constraints, typically called "capital requirements". In a scenario where the regulatory authority can perfectly assess the riskiness of a bank's credit portfolio, it can make use of risk-sensitive leverage constraints

[3]Note that our rationale for a maximum leverage ratio, that is, forcing banks to keep enough skin in the game to guarantee a certain level of aggregate monitoring, is different from the systemic-risk mitigation rationale for such a constraint, as brought forward by Morris and Shin (2008).

and replicate the first-best allocation in the MC economy. The reason is that the regulator will threaten banks with a tight leverage constraint if their credit portfolio turns out to be high-risk, which is the case if they shirk monitoring, but will set a loose leverage constraint for low-risk banks, i.e., banks who monitor. As a consequence, all banks opt for monitoring and capital is allocated efficiently.

2 Broader Implications and Literature

Our analysis also allows to assess whether the standard LF approach, which is typically used in macroeconomic modeling, is a valid shortcut for modeling the banking sectors' main role within the economy. In contrast to Faure and Gersbach (2022), who show that the LF economy and the MC economy produce equivalent outcomes when considering an environment without moral hazard at bank level, our findings show that this result does not carry over to a setting with heterogeneous banks and financial frictions. An inefficiently low allocation of capital to bank-dependent firms, due to bank-level moral hazard, turns out to be less of a worry in our actual monetary system with bank money creation than what the LF approach would suggest.[4] Hence, our results imply that while in many circumstances, using the LF approach may be sufficient, it is not adequate in other circumstances. In particular, if we want to understand the functioning, optimal regulation and policy-making in our current monetary system, one should use the MC approach—and many bells and whistles can be added to the model in future research. We expect that accounting for the dual role of banks as loan providers and money creators will become more important as this area of research expands.[5]

We also show that while the MC economy produces higher aggregate output, it is more fragile than the LF economy. This is because the MC economy depends on the regulator correctly setting the leverage constraint. If this is not the case, welfare in the MC economy can be lower than in the LF economy.

The practice of money and loan creation by commercial banks has a long

[4]A parallel argument was made by Jakab and Kumhof (2019) within a DSGE approach.

[5]This may also be important in education. As emphasized in an article in *The Economist*, we should continuously review whether the simplified models we teach depict reality adequately. See "Efforts to modernise economics teaching are gathering steam", *The Economist*, March 18th 2021 edition, `https://www.economist.com/finance-and-economics/2021/03/20/efforts-to-modernise-economics-teaching-are-gathering-steam`.

history and has been subject to enduring analyses and debates (Macleod, 1866; Wicksell, 1907; Hahn, 1920; Keynes, 1931; Schumpeter, 1954; Gurley and Shaw, 1960; Tobin, 1963; McLeay et al., 2014; Donaldson et al., 2018). In modern times, the money banks create is a claim on fiat money which is created by the central bank. Different modeling approaches are pursued and applied to capture this (Skeie, 2008; Jakab and Kumhof, 2019; Wang, 2019; Bolton et al., 2020; Faure and Gersbach, 2021; Piazzesi et al., 2021; Wang, 2021; Li and Li, 2021; Parlour et al., 2022).[6] In this paper, we provide a rationale why our current monetary system, in which banks have the privilege to create private money as claims on public fiat money, is advantageous when there is unobservable heterogeneity among banks.

Our paper involves a simple set of reasons why bank deposits as claims on fiat money have a positive value as a medium of exchange. First, firms can only acquire investment goods from households if they obtain loans from banks in the form of bank deposits. Second, households accept the firms' bank deposits, since they can later use them to acquire the consumption goods produced by firms. Third, firms provide the consumption goods in return for the households' bank deposits because they need to repay their bank loans. Finally, banks repay their loans from the CB, since they face large penalties in case of default. Hence, all money that was created at the beginning of the economy is destroyed at the end: bank money is destroyed when firms repay their bank loans, CB money is destroyed when banks repay the CB. Our paper is thus a variant of theories that examine under which circumstances fiat money can have positive value in finite-horizon settings (see models and discussions, for instance in Shubik and Wilson, 1977; Dubey and Geanakoplos, 1992, 2003a,b, 2006; Shapley and Shubik, 1977; Shubik and Tsomocos, 1992; Tsomocos, 2003; Bloise and Polemarchakis, 2006; Goodhart et al., 2006).[7]

The paper is organized as follows. Section 3 introduces the LF economy and solves for equilibrium. Section 4 does the same for the MC economy, taking the regulatory leverage constraint as given. Section 5 derives the optimal leverage constraint in the MC economy and compares the resulting allocations to those

[6] A parallel literature has examined the properties of monetary systems when banks issue banknotes instead of deposits (e.g., Gersbach, 1998; Cavalcanti and Wallace, 1999).

[7] See Huber et al. (2014) for a summary of the reasons why the value of fiat money can be positive in finite and infinite horizon models.

in an LF economy. Section 6 illustrates how bank-risk assessments, combined with risk-sensitive leverage constraints, can further improve outcomes in the MC economy or even achieve first-best. Section 7 concludes. Proofs are relegated to the Appendix.

3 Loanable Funds

3.1 The model

First, we introduce the model in the LF setting. Consider a two-period economy ($t = 1, 2$) with two types of goods: a capital good and a consumption good. The capital good is used as the sole input factor in firms' production of the consumption good. Returns are expressed in terms of the consumption good. There are three types of risk-neutral agents: households, bankers and entrepreneurs. All agents are price-takers.

Entrepreneurs run firms but need external financing to realize their projects. In $t = 1$, households provide capital goods to firms, either through direct financing in the bond market or through indirect financing, which requires intermediation by banks. In $t = 2$, firms produce and consumption takes place. The total endowment of the capital good in period 1 is normalized to one.

Let us next describe the agents' roles in more detail.

Entrepreneurs. Entrepreneurial activity takes place in two separated productive sectors, which differ in production technologies and financing options. There is a continuum of entrepreneurs. Entrepreneurs run firms and have no endowment. Firms in the first sector (the bank-dependent sector, henceforth, "BS") can only acquire indirect financing via banks. We do not explicitly model why this is the case, but one could think of a firm-level moral hazard problem that requires these firms to obtain external governance from an intermediary. Firms in the BS have access to a risky production technology that yields a constant gross return to scale sR_B, where

$$s = \begin{cases} 1 & \text{if production is successful,} \\ 0 & \text{if production fails.} \end{cases}$$

The probability of success depends on banks' monitoring efforts (see below). The aggregate amount of capital lent to firms in the BS is denoted by K_B.

Firms within the second productive sector (the frictionless sector, henceforth, "FS") have sound internal governance and thus have access to direct financing from households through the bond market. The production technology in the FS is characterized by diminishing returns to scale at the aggregate level. There is no productive uncertainty in the frictionless sector. If we denote the total amount of capital given to firms in the FS by K_F, output in terms of the consumption good is given by $g(K_F)$, where $g'(K_F) > 0$, $g''(K_F) < 0$ and $\lim_{K_F \to 0} g'(K_F) = \infty$. Profit maximization entails that households' gross return R_F per unit of capital invested into the FS is given by $R_F = g'(K_F)$.

Bankers. There is a continuum of bankers indexed by $b \in [\underline{b}, \overline{b}]$. Each banker owns and runs a bank and each bank is endowed with e units of the capital good, i.e., e denotes a bank's equity, where $0 < e < (\overline{b} - \underline{b})^{-1}$. Aggregate bank equity is $E = (\overline{b} - \underline{b})e$ and thus $0 < E < 1$. Each bank b takes household deposits d_b and promises a per unit repayment R_D in case of success. Hence, the deposit gross rate is sR_D.[8] The bank uses acquired household fundings, together with its own equity, to lend an amount k_b $(= d_b + e)$ at gross rate sR_L to firms within the BS. Constant returns to scale imply zero profits for BS firms. Hence, $R_L = R_B$. Note that all returns are stated as gross returns. For the sake of brevity, we will often simply use the term "return".

Each bank b faces a monitoring decision $\gamma_b \in \{0, 1\}$: it either diligently engages in loan monitoring ($\gamma_b = 1$) or shirks such efforts ($\gamma_b = 0$). If a bank monitors, its borrowing firms' probability of success in production is given by π (with $0 < \pi < 1$). If a bank shirks monitoring, this probability decreases to $\pi - \Delta$ (with $0 < \Delta < \pi$), but the banker enjoys a private benefit b (> 0) per unit of lending. Since banks differ with respect to b, there is heterogeneity among banks regarding their private benefits from shirking and hence regarding their incentives for moral hazard behavior.

Households. There is a continuum of identical households (HHs), so that we can

[8] In case of failure, i.e., for $s = 0$, households' deposits are lost and the households face a gross rate of return equal to zero. An introduction of partial deposit insurance would not fundamentally change our results.

focus on a representative household. The aggregate amount of capital households are endowed with is $1 - E$. Households maximize consumption in period $t = 2$ by optimally allocating their capital goods between the two productive sectors, i.e., by optimally providing capital either to the FS by buying bonds or to the BS by investing in bank deposits. We will focus on "interior" allocations, where households provide positive amounts of capital to both sectors. In this case, households' expected returns from bonds and deposits have to equalize.

3.2 First-best

Before solving for the competitive equilibrium of our economy, we characterize the properties of the first-best. Throughout the paper, we assume that loan monitoring by banks is economically efficient.

Assumption 1 (Economically efficient monitoring technology)
Let $\Delta R_B \geq \bar{b}$.

Assumption 1 states that the additional expected output created if Bank $\bar{b}$ monitors compared to if it does not, given by $\pi R_B k_{\bar{b}} - (\pi - \Delta) R_B k_{\bar{b}}$, exceeds the bank's private benefits $\bar{b} k_{\bar{b}}$ from non-monitoring. As $b \in [\underline{b}, \bar{b}]$, this (strictly) applies also for all other banks b.

Welfare criterion. Since all agents are risk-neutral, we take expected aggregate output as the welfare criterion for our economy. This specification neglects bankers' private benefits, which, however, does not affect our main findings. Assumption 1 implies that the first-best requires monitoring efforts by all banks (i.e., $\gamma_b = 1$ for all b), irrespective of whether we account for the bankers' private benefits or not. With regard to a comparison of the LF and MC economies, which this paper ultimately aims for, an extended welfare criterion that would take bankers' private benefits into account would only reinforce our results.[9]

In the following Proposition, we characterize the first-best. The first-best values for K_B and K_F are denoted by K_B^{FB} and K_F^{FB}.

[9] This is due to the fact that, as we will show, equilibrium in the LF economy entails monitoring by all banks, while there are also non-monitoring banks in the MC economy. Even when neglecting bankers' private benefits from non-monitoring, welfare is higher in the MC economy (cf. Proposition 6). Thus it would certainly also be higher if we would take these private benefits into account.

Proposition 1 (First-best)
In first-best, $\gamma_b = 1$ for all b and thus the success probability of bank-dependent firms' projects is π. Capital is allocated according to $K_F^{FB} = (g')^{-1}(\pi R_B)$ and $K_B^{FB} = 1 - K_F^{FB}$.

The first-best values K_B^{FB} and K_F^{FB} are derived from the fact that capital is allocated efficiently between the two productive sectors and hence the marginal returns equalize, i.e., $\pi R_B = g'(K_F^{FB})$, and that all capital is used, i.e. $K_B^{FB} + K_F^{FB} = 1$.

3.3 Equilibrium

We now turn to analyzing the behavior of households, banks and firms in a competitive equilibrium of the LF economy. Since households cannot distinguish between bank types, i.e., b is unobservable, all banks receive equal amounts of deposits $d_b = d$, which implies that $k_b = k$ is constant across banks. We construct an equilibrium in which all banks monitor. Given a loan amount k, Bank $\bar{b}$ monitors if its expected additional profits when monitoring exceed its private benefits from shirking:[10]

$$\Delta \left[R_B k - R_D(k - e) \right] \geq \bar{b} k. \tag{1}$$

Rewriting this condition yields

$$k(R_D - R_B + \frac{\bar{b}}{\Delta}) \leq e R_D. \tag{2}$$

With Assumption 1 and $k = d + e$, a bank funding its loans solely through equity (i.e., $d = 0$) would always opt for diligent loan monitoring. With positive levels of household funding (i.e., $d > 0$), this is not necessarily the case. Condition (2) gives the maximum incentive-compatible amount of capital that households can provide to Bank $\bar{b}$. If household deposits d and thereby the loan amount k would exceed the value for which Condition (2) holds with equality, Bank $\bar{b}$'s additional expected profits from monitoring would fall short of its private benefits from shirking. In other words, the bank would not have enough skin in the

[10] If the bank with the highest private benefits from shirking, i.e., Bank $\bar{b}$, monitors, then of course all other banks monitor as well.

game to behave diligently. We call Condition (2) the *incentive constraint*, which households have to respect if they want all banks to monitor.

Households provide funding to banks only if the expected return on deposits is not lower than the return R_F from bonds issued by firms in the FS. Given that the incentive constraint holds, households' expected return on deposits is given by πR_D. Hence, households' *participation constraint* for investment in the BS through deposits is given by

$$\pi R_D \geq R_F. \tag{3}$$

In an "interior" allocation in which households provide positive amounts of capital to both productive sectors, Condition (3) must be satisfied with equality. As we want to focus on such cases, we make the following assumption:

Assumption 2 (Bank lending exceeds bank equity)
Let $g'(1-E) < (\pi - \Delta)R_B$.

Assumption 2 states that the marginal product of capital in the FS falls short of the (expected) marginal product of capital in the BS, as long as the total amount of capital deployed to the BS does not exceed aggregate bank equity E. Thereby, the assumption ensures that the amount of capital flowing to the BS exceeds aggregate bank equity E. This is independent of whether banks monitor or not, since from Assumption 2 immediately follows also $g'(1-E) < \pi R_B$.

Together with the Inada condition $\lim_{K_F \to 0} g'(K_F) = \infty$, Assumption 2 implies that households provide positive amounts of capital to both productive sectors. It follows that the expected returns from holding deposits or bonds equalize in equilibrium, i.e., Condition (3) is satisfied with equality. Then, substituting Condition (3) into Condition (2), banks maximize profits by solving the following constrained optimization problem:

$$\begin{aligned} &\max_k \ \pi R_B k - R_F(k-e), \\ &s.t. \ k(R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}) \leq e R_F. \end{aligned} \tag{4}$$

Note that if all households respect the incentive constraint (2) and thus all banks monitor, a single (price-taking) household has no incentive to deviate by offering an amount of deposits that exceeds the maximum incentive-compatible one. The promised return R_D ($= R_F/\pi$) on risky deposits is such that the household is indifferent between *monitored* investment in the BS and bonds from the FS. If the household violates Condition (2), it is at risk of depositing at a bank which then no longer monitors. The given return R_F/π, however, does not offer a compensation for that risk. Hence, the household is better off adhering to Condition (2) and investing its remaining capital into the FS at rate R_F.

In the following proposition, we first characterize the "all-monitor" equilibrium of the LF economy. The equilibrium values for R_F and K_B are denoted by R_F^{LF} and K_B^{LF}.

Proposition 2 (Deficient bank-funding in the LF economy)
There is a competitive equilibrium with $\gamma_b = 1$ for all b. If bank equity is scarce, i.e., for

$$E < \bar{E}^{LF} := \frac{\bar{b}\,[1-(g')^{-1}(\pi R_B)]}{\Delta R_B}, \tag{5}$$

the constraint in the maximization problem in (4) is binding. Then, the equilibrium return R_F^{LF} is given by the solution to

$$R_F = g'\left\{1 - \frac{eR_F(\bar{b}-\underline{b})}{R_F - \pi R_B + \frac{\pi\bar{b}}{\Delta}}\right\}$$

and satisfies $R_F^{LF} < \pi R_B$. It follows that there is underinvestment in the bank-dependent sector, i.e., $K_B^{LF} < K_B^{FB}$.

The proof is in Appendix A. Proposition 2 states that if bank equity is scarce, households are constrained in the incentive-compatible amount of deposits they can provide to banks. As a consequence, the equilibrium spread between the return on investment in BS and FS firms is positive: $\pi R_B > R_F^{LF}$. It follows that aggregate bank lending is inefficiently low when compared to the first-best. Also note that for given aggregate bank equity E, Condition (5) is more likely to hold if financial frictions are large, i.e., if $\bar{b}$ is large. In case that banks hold enough equity such that Condition (5) is violated, the constraint in the maximization problem in (4) is non-binding and the all-monitor equilibrium of the LF economy coincides with the first-best.

Uniqueness of equilibrium. Finally, we observe that there can be no other equilibria that involve bank monitoring besides the all-monitor equilibrium established above. To see this, consider a scenario where households provide an amount of bank deposits such that the incentive constraint is met only for banks with $b \leq \hat{b}$, where $\hat{b} \in (\underline{b}, \overline{b})$, and thus only a fraction of banks monitors. Denote the resulting average success probability of a BS investment by μ $(< \pi)$. To qualify as an equilibrium, it has to be $\mu R_D = R_F$. But then an individual (price-taking) household would have an incentive to deviate by reducing its amount of deposits, since this would increase the probability that its depositing bank monitors. A given return of $R_D = R_F/\mu$ would then imply an overcompensation for the risk related to BS investment. Capital that was freed through the reduction of bank deposits could be invested at rate R_F, which is exactly the rate for which the household was indifferent between deposits and FS bonds in the first place. Hence, reducing investment in banks dominates the original investment.[11]

4 Money Creation

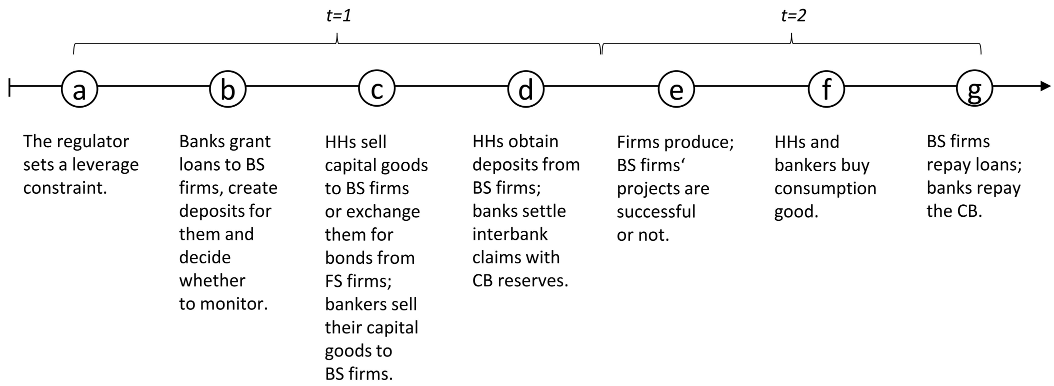
4.1 The model

We now turn to the MC economy. First, the monetary and regulatory framework of the two-tier monetary architecture is presented. The elements of the model related to the real side of the economy are the same as in the LF setting.

Money in the MC economy comes in two forms, bank deposits and CB reserves. Bank deposits are created when banks grant loans. They are used for payment between non-bank entities. Reserves are held by banks at the CB. They are used to settle interbank transactions. In contrast to the LF economy, banks do not only act as simple intermediaries which collect household deposits and subsequently lend these to firms. Instead, banks create new deposits when they make loans: the amount a firm borrows simultaneously appears on its bank's balance sheet

[11]There are two caveats to make here. First, we assume that a single household is pivotal to turn one bank from non-monitoring into monitoring by reducing its investment. As long as households invest into a finite number of banks, this is satisfied. Second, there might also be an "all-shirk" equilibrium, i.e., an equilibrium where all banks shirk. This could happen in the extreme case that households have spread their investments across many banks and provide so much bank capital that it is impossible for an individual household to meet any bank's incentive constraint by reducing only its own amount of deposits.

Figure 1: Timeline for the MC economy



as a deposit. Firms use these deposits to buy the capital good from households (and from banks). Interbank transactions, which arise from the fact that agents may hold accounts at different banks, are settled by reserves. Banks can obtain reserves either by borrowing in a competitive interbank market or by taking a loan from the CB. Households use the deposits they receive from selling the capital good to buy the consumption good. Firms use the deposits they receive from selling the consumption good to pay back their loans. Finally, banks repay their interbank loans and their loans from the CB. At the beginning, the regulator sets a leverage constraint. Figure 1 summarizes the sequence of events. We next describe the model in formal terms. For the moment, we neglect the regulatory leverage constraint, which will be introduced in the next subsection.

Monetary framework. With money in the model, we now have to distinguish between real and nominal variables. To be clear in this regard, we use bold fonts to indicate real variables and normal fonts to indicate nominal variables. The BS prices of the capital good and the consumption good are denoted by p_I and p_C, respectively. Bank loans to firms are stated in nominal terms as well. They are denoted by l_b on the individual level and by L_B $(= \int_{\underline{b}}^{\overline{b}} l_b \, \mathrm{d}b)$ on the aggregate level. A loan of amount l_b buys a firm l_b/p_I units of the capital good. Households sell their capital goods to firms in the FS or BS. The aggregate amount of capital deployed to the FS is denoted by $\boldsymbol{K_F}$ and yields households a risk-free real return $\boldsymbol{R_F K_F}$ (in terms of the consumption good). Households sell their remaining capital goods $(\boldsymbol{1} - \boldsymbol{K_F} - \boldsymbol{E})$ to BS firms and are credited with deposits of nominal value $(\mathbf{1} - \boldsymbol{K_F} - \boldsymbol{E})p_I$. Bankers sell their capital goods to BS firms and receive $\boldsymbol{e}p_I$ deposits each in return. As these are claims on themselves, $\boldsymbol{e}p_I$ is the nominal amount of each bank's equity.

Financial frictions. For our outcomes in the MC and LF economies to be comparable, we make sure that we consider exactly the same frictions in both settings. In the LF economy, bankers' private benefits from skipping monitoring efforts were related to *real* lending $\boldsymbol{k_b}$. In the MC economy, we relate bankers' private benefits to *nominal* lending l_b. Therefore, we have to take the price ratio p_C/p_I into account, i.e., we assume that a non-monitoring banker enjoys (nominal) private benefits $(p_C/p_I)\boldsymbol{b}$ per unit of lending l_b.

CB policy rate. We denote the gross rate for borrowing from (or depositing at) the CB by R_{CB}.[12] Assume that there is a competitive interbank market and that banks cannot discriminate between deposits owned by households and deposits owned by other banks. Then, a simple no-arbitrage argument establishes Lemma 1.

Lemma 1 (Deposit rate equals CB policy rate)
The deposit rate equals the CB policy rate:

$$R_D = R_{CB}. \tag{6}$$

Proof. Our assumptions with regard to the interbank market immediately imply that the gross interbank rate has to equal households' deposit rate R_D. Then, for $R_{CB} > R_D$, all banks would borrow from other banks in order to hold reserves at the CB. This cannot be an equilibrium. On the other hand, for $R_{CB} < R_D$ every bank would want to take a loan of infinite amount from the CB and subsequently lend the acquired funds to other banks to generate profits. This can also be no equilibrium.

For $R_D = R_{CB}$, an individual bank is indifferent between participating in the interbank market and transacting with the CB. Without loss of generality, we assume that it chooses the latter.

Interbank transactions and CB reserves. After the capital good has been

[12]We note that the assumption of a zero spread between central bank rates simplifies the analysis but is not vital to our results.

sold to firms, each bank b faces one of the following two scenarios: (i) $l_b < d + \boldsymbol{e}p_I$, i.e., deposit inflows exceed deposit outflows, or (ii) $l_b > d + \boldsymbol{e}p_I$, i.e., deposit outflows exceed deposit inflows. Deposit outflows are given by l_b, since the amount of loans bank b grants to firms is credited as deposits to the firms' bank accounts and these deposits leave the bank when the firms acquire the capital good from households with accounts at different banks. Analogously, each bank experiences an inflow of deposits d when households with accounts at the bank, as well as the bank itself, sell their capital goods to firms with accounts at different banks.[13] In case (i), the bank holds reserves in the amount of the net inflow of deposits $d + \boldsymbol{e}p_I - l_b$ at the CB. In case (ii), the bank has to borrow reserves in the amount of the net outflow of deposits $l_b - d - \boldsymbol{e}p_I$ from the CB in order to be able to cover its interbank liabilities. Obviously, the higher a bank's lending volume l_b the higher the amount of reserves the bank has to acquire in order to support it.

Profit function of an individual bank. Bank b's (nominal) profit function when monitoring Π_π is given by

$$\Pi_\pi = \pi[R_L l_b - R_D d - R_{CB}(l_b - d - \boldsymbol{e}p_I)].$$

With probability π the investment is successful and the bank receives a return R_L on its loans, pays households a return R_D on their deposits and, depending on whether scenario (i) or (ii) applies, receives or pays a rate R_{CB} on its CB reserves. With probability $1 - \pi$ the investment fails and profits are zero. Using Equation (6), we can simplify the bank's profit function to

$$\Pi_\pi = \pi[R_L l_b - R_D(l_b - \boldsymbol{e}p_I)]. \tag{7}$$

If bank b shirks monitoring, it enjoys private benefits but the success probability of its investment drops to $\pi - \Delta$. In this case, its (nominal) profit function Π_Δ is given by

$$\Pi_\Delta = (\pi - \Delta)[R_L l_b - R_D(l_b - \boldsymbol{e}p_I)] + \boldsymbol{b}\frac{p_C}{p_I} l_b. \tag{8}$$

No default against the CB. When households use their deposits to buy the

[13] We assume that households distribute deposits evenly across banks and thus the amount of household deposits at each bank does not depend on the individual bank type b.

consumption good, bank b experiences an outflow of deposits equal to sR_Dd. Firms use the funds they receive from households to pay back their loans and bank b receives an amount sR_Ll_b. Independent of whether it monitors or not, the bank can repay its CB loans, whenever $s = 1$ and

$$R_Ll_b - R_Dd \geq (l_b - d - \boldsymbol{e}p_I)R_{CB}.$$

With $R_{CB} = R_D$, this comes down to

$$l_b(R_D - R_L) \leq R_D\boldsymbol{e}p_I. \tag{9}$$

We make the following assumption:

Assumption 3 (No default against the CB)
Each bank b respects Condition (9) when deciding on its loan volume l_b.

We note that as long as $R_L \geq R_D$, Condition (9) always holds.[14]

4.2 Equilibrium considerations

Banks' monitoring decision. Banks choose the amount of lending they grant to firms and decide whether to exert effort in monitoring or not. Given l_b, bank b monitors if its expected additional profits when monitoring exceed its private benefits from shirking. From Equations (7) and (8), this is the case exactly if

$$\Delta[R_Ll_b - R_D(l_b - \boldsymbol{e}p_I)] \geq \boldsymbol{b}\frac{p_C}{p_I}l_b. \tag{10}$$

Solving for $\boldsymbol{b}$ yields

$$\boldsymbol{b} \leq \hat{\boldsymbol{b}} := \frac{p_I}{p_C}\left[\Delta(R_L - R_D) + \Delta\frac{R_D\boldsymbol{e}p_I}{l_b}\right]. \tag{11}$$

Intuitively, only banks with sufficiently low opportunity costs of monitoring $\boldsymbol{b}$ decide to do so. We define $\hat{\boldsymbol{b}}$ as the threshold value which divides the continuum

[14] If a bank monitors, $R_L \geq R_D$ is necessary to provide it with an incentive to lend positive amounts anyway. Non-monitoring banks, however, may have an incentive to lend even for $R_L < R_D$, as they also enjoy private benefits $\boldsymbol{b}(p_C/p_I)l_b$. Hence, Condition (9) may not necessarily hold for non-monitoring banks.

of banks into a monitoring part $[\underline{\boldsymbol{b}}, \hat{\boldsymbol{b}}]$ and a non-monitoring part $(\hat{\boldsymbol{b}}, \overline{\boldsymbol{b}}]$. Whether $\hat{\boldsymbol{b}}$ will be indeed interior to $[\underline{\boldsymbol{b}}, \overline{\boldsymbol{b}}]$ or whether extreme cases occur—all or no banks monitor—depends on prices, interest rates and loan volumes.

The regulator can limit banks' lending volumes via a leverage constraint. To see that such a constraint affects banks' monitoring decision, we rewrite Inequality (10) as

$$\underbrace{\frac{l_b}{\boldsymbol{e} p_I}}_{=:\alpha_b} \left[\boldsymbol{b}\frac{p_C}{p_I} - \Delta(R_L - R_D)\right] \leq \Delta R_D,$$

where α_b denotes bank b's leverage ratio (i.e., loans over equity). For a leverage ratio equal to one, Assumption 1 implies that all banks monitor. For leverage ratios greater than one, this is not necessarily the case. But even then, banks with $\boldsymbol{b} \leq (p_I/p_C)\Delta(R_L - R_D)$ always monitor. For $\boldsymbol{b} > (p_I/p_C)\Delta(R_L - R_D)$, however, bank b only monitors if its leverage ratio is not too high so that it has enough skin in the game:

$$\alpha_b \leq \frac{\Delta R_D}{\boldsymbol{b}\frac{p_C}{p_I} - \Delta(R_L - R_D)}. \tag{12}$$

Banks' lending decision. As long as $R_L > R_D$ $(= R_{CB})$, a bank's profit function is linearly increasing in l_b. This holds true both in case that the bank monitors and its profits are given by Π_π according to Equation (7) and in case that the bank shirks and its profits are given by Π_Δ according to Equation (8).[15] Hence, as long as the lending rate exceeds the deposit rate, banks lend as much as possible and they are constrained only by a regulatory leverage constraint, which we introduce next.

Regulatory leverage constraint. The regulator sets a leverage constraint α, which simply specifies an upper limit on banks' leverage ratios α_b. Since the regulator cannot distinguish between bank types, it has to choose a universal constraint. Therefore, it faces a trade-off. Setting a tight leverage constraint ensures that Condition (12) holds for most banks and hence also those with high $\boldsymbol{b}$ monitor, but strongly constrains bank lending and thereby leads to lower-than-

[15] As banks' private benefits from shirking monitoring efforts scale with l_b, a non-monitoring bank's profits are linearly increasing in l_b even if $R_L < R_D$ but still $(\pi - \Delta)(R_L - R_D) + \boldsymbol{b}(p_C/p_I) > 0$.

optimal lending levels for banks with low $\boldsymbol{b}$ and thus inefficiently low volumes of capital provided to the BS. On the other hand, setting a loose leverage constraint increases banks' lending capacity and allows bank-dependent firms to acquire funding more easily thereby promoting a more efficient allocation of capital, but also implies that Condition (12) is violated for most banks and hence aggregate monitoring activity is low.

The regulator decides on the leverage constraint at the beginning of $t = 1$, anticipating agents' equilibrium reactions. We take α as given in the equilibrium analysis and solve for its output-maximizing value in Section 5. Importantly, we restrict attention to regulatory leverage constraints that are binding for all banks b, i.e., to values of α for which the banks' lending rate R_L still exceeds the deposit rate R_D, when all banks b leverage up to $\alpha_b = \alpha$. We denote by $\hat{\alpha}$ (> 1) the threshold value for α, below which the leverage constraint is binding for all banks. Hence, we restrict attention to $\alpha \in [1, \hat{\alpha})$. Later on, Lemma 3 will determine $\hat{\alpha}$ and Corollary 2 will show that the optimal leverage constraint lies within this interval.

Households' investment decision. In an interior equilibrium where risk-neutral households hold bonds and deposits, the expected real returns from bonds and deposits have to equalize. Investing one unit of the capital good into the FS (via bonds) gives a certain return $\boldsymbol{R_F}$. Investing one unit of the capital good in the BS (via bank deposits) yields an expected real return $q(p_I R_D / p_C)$, where q gives the average success probability of BS investments:

$$q = \begin{cases} \pi & \text{for } \hat{\boldsymbol{b}} \geq \overline{\boldsymbol{b}}, \\ \mu := \pi - \frac{\overline{\boldsymbol{b}} - \hat{\boldsymbol{b}}}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \Delta & \text{for } \underline{\boldsymbol{b}} < \hat{\boldsymbol{b}} < \overline{\boldsymbol{b}}, \\ \pi - \Delta & \text{for } \hat{\boldsymbol{b}} \leq \underline{\boldsymbol{b}}. \end{cases} \tag{13}$$

For $\hat{\boldsymbol{b}} \geq \overline{\boldsymbol{b}}$ all banks monitor, for $\hat{\boldsymbol{b}} \leq \underline{\boldsymbol{b}}$ all banks shirk. For values of $\hat{\boldsymbol{b}}$ in between, some banks monitor while others do not. For the relationship between $\boldsymbol{R_F}$ and R_D, we obtain

$$\boldsymbol{R_F} = q \frac{p_I R_D}{p_C}. \tag{14}$$

BS market clearing. Capital and consumption goods markets in the BS clear.

For the capital goods market, this implies

$$\boldsymbol{K_B} = \frac{L_B}{p_I}. \tag{15}$$

For the consumption goods market, it implies:

$$\underbrace{qp_C\boldsymbol{R_B}\boldsymbol{K_B}}_{exp.\ BS\ firm\ supply} = \underbrace{(L_B - \boldsymbol{E}p_I)qR_D}_{exp.\ household\ demand} + \underbrace{(L_B - \boldsymbol{E}p_I)q(R_L - R_D) + q\boldsymbol{E}p_I R_L}_{exp.\ bank\ demand}. \tag{16}$$

Remember that firms in the BS make zero profits, i.e., banks extract the entire surplus. Simplifying Equation (16) by using Equation (15) yields

$$\begin{aligned} p_C\boldsymbol{R_B}\boldsymbol{K_B} &= (p_I\boldsymbol{K_B} - \boldsymbol{E}p_I)R_L + \boldsymbol{E}p_I R_L \\ \boldsymbol{R_B} &= \frac{p_I R_L}{p_C}. \end{aligned} \tag{17}$$

A competitive equilibrium of the MC economy is then defined as follows.

Definition 1 (Equilibrium of the MC economy)
Given the CB policy rate R_{CB} and a regulatory leverage constraint $\alpha \in [1, \hat{\alpha})$, a competitive equilibrium is a BS capital to goods price ratio p_I/p_C, loan and deposit rates R_L and R_D, a FS capital price $\boldsymbol{R_F}$, individual bank monitoring decisions γ_b and lending plans l_b, such that

(i) given p_I/p_C, R_L, R_D and $\boldsymbol{R_F}$, individual lending plans l_b maximize the expected profit of each bank subject to the leverage constraint α;

(ii) given p_I/p_C, R_L, R_D, $\boldsymbol{R_F}$ and l_b, each bank optimally decides whether to monitor or not;

(iii) given p_I/p_C, R_D and $\boldsymbol{R_F}$, households optimally invest their capital goods;

(iv) aggregate demand for capital equals aggregate supply:

$$\boldsymbol{K_B} = 1 - \boldsymbol{K_F} \quad or \quad \boldsymbol{R_F} = g'(1 - \boldsymbol{K_B});$$

(v) capital and consumption goods markets in the BS clear.

By Lemma (1), the CB rate R_{CB} immediately pins down R_D. Since $\alpha \in [1, \hat{\alpha})$, *(i)* implies $l_b/p_I = \alpha \boldsymbol{e}$ for all banks b, independent from their monitoring decisions. Bank monitoring decisions γ_b and the equilibrium threshold value $\hat{\boldsymbol{b}}$ result from *(ii)*, and they determine the average success probability q. Conditions *(iii)–(v)* then determine $\boldsymbol{R_F}$, p_I/p_C and R_L.[16] We explicitly solve for the equilibrium values in Subsection 4.4.

4.3 Special case: tight leverage constraint (all monitor)

With regard to the regulatory leverage constraint α, let us first take a look at the extreme case where the regulator sets α such that $\gamma_b = 1$ for all b. From Condition (12), this constraint would be given by

$$\alpha = \frac{\Delta R_D}{\bar{\boldsymbol{b}}\frac{p_C}{p_I} - \Delta(R_L - R_D)}. \tag{18}$$

Using Equations (17) and (14) with $q = \pi$, Equation (18) becomes

$$\alpha = \frac{\boldsymbol{R_F}}{\boldsymbol{R_F} - \pi \boldsymbol{R_B} + \frac{\pi \bar{\boldsymbol{b}}}{\Delta}}. \tag{19}$$

The leverage constraint sets an upper limit for bank lending:

$$\underbrace{\frac{l_b}{p_I}}_{\equiv \boldsymbol{k_b}} \leq \frac{\boldsymbol{e}\boldsymbol{R_F}}{\boldsymbol{R_F} - \pi \boldsymbol{R_B} + \frac{\pi \bar{\boldsymbol{b}}}{\Delta}}. \tag{20}$$

We can state the following proposition.

Proposition 3 (Equilibrium of the MC economy, tight regulation) *Let the regulatory leverage constraint be given by Equation (19), so that $\gamma_b = 1$ for all b. Then, the MC economy yields the same economic outcomes as the LF economy.*

As Inequality (20) corresponds to Inequality (A.1) in Appendix A, the proof is straightforward. The question is the following: can the regulator do better by

[16] We note that as usual in a monetary economy, the "initial" price p_I is not determinate (e.g., Benigno and Nisticò, 2022). Without loss of generality, we could normalize $p_I = 1$.

allowing for higher leverage ratios, which, however, imply that some banks won't monitor?

4.4 The general case

Consider now the general case where the regulator may set a leverage constraint α that implies positive fractions of both monitoring and non-monitoring banks. We continue to restrict attention to $\alpha \in [1, \hat{\alpha})$, which implies $R_L > R_D$ in equilibrium. By Equations (14) and (17), this in turn implies $q\boldsymbol{R_B} > \boldsymbol{R_F}$, that is, the leverage constraint restricts the amount of lending to the BS, such that the (expected) marginal product of capital in the BS exceeds the marginal product of capital in the FS.

From Condition (9), we know that $R_L > R_D$ also implies that banks can always repay the CB (if $s = 1$). Hence, for any regulatory leverage constraint $\alpha \in [1, \hat{\alpha})$, it holds that

$$\begin{aligned} l_b &= \alpha \boldsymbol{e} p_I, \\ L_B &= \alpha \boldsymbol{E} p_I, \\ \boldsymbol{K_B} &= \alpha \boldsymbol{E}. \end{aligned} \tag{21}$$

With $g'(\boldsymbol{K_F}) = \boldsymbol{R_F}$, it follows that

$$\boldsymbol{R_F} = g'(1 - \alpha \boldsymbol{E}). \tag{22}$$

The threshold value $\hat{\boldsymbol{b}}$ for monitoring banks is then given by Expression (11), making use of Equations (6) and (17):

$$\begin{aligned} \hat{\boldsymbol{b}} &= \frac{p_I}{p_C} \left[\Delta(R_L - R_D) + \frac{\Delta}{\alpha} R_D \right] \\ &= \Delta \boldsymbol{R_B} - \frac{p_I}{p_C} \left(1 - \frac{1}{\alpha} \right) \Delta R_{CB}. \end{aligned} \tag{23}$$

Equilibrium BS price ratio. Substituting $\hat{\boldsymbol{b}}$ into Expression (13) yields q. Substituting q into Equation (14) yields the equilibrium BS price ratio. For $q = \pi$, we obtain $p_I/p_C = \boldsymbol{R_F}/(\pi R_{CB})$. For $q = \pi - \Delta$, we obtain $p_I/p_C = \boldsymbol{R_F}/[(\pi -$

$\Delta)R_{CB}]$.[17] For $q = \mu$, we obtain

$$\begin{aligned}\boldsymbol{R_F} &= \mu \frac{p_I}{p_C} R_{CB} \\ &= \frac{p_I}{p_C} \underbrace{R_{CB}\mu_1}_{=:B>0} - \left(\frac{p_I}{p_C}\right)^2 \underbrace{R_{CB}^2 \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \left(1 - \frac{1}{\alpha}\right)}_{=:A>0},\end{aligned} \tag{24}$$

where

$$\mu_1 := \pi - \frac{\overline{\boldsymbol{b}} - \Delta \boldsymbol{R_B}}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \Delta. \tag{25}$$

From Expression (13) and Equation (23), μ_1 corresponds to μ evaluated at $\alpha = 1$. With $\boldsymbol{R_F}$ given by Equation (22), Equation (24) implicitly determines the equilibrium price ratio. For $\alpha = 1$, it is $A = 0$ and we obtain $p_I/p_C = \boldsymbol{R_F}/(\mu_1 R_{CB})$. For $\alpha > 1$, Equation (24) is a quadratic equation in p_I/p_C, which we can solve explicitly:

$$\frac{p_I}{p_C} = \frac{1}{2A}\left(B \pm \sqrt{B^2 - 4A\boldsymbol{R_F}}\right), \tag{26}$$

with A and B as indicated in Equation (24).

Existence. A solution to Equation (24) exists, if the term under the square root function in Equation (26) is non-negative. Denote the value of α that solves $B^2 = 4A\boldsymbol{R_F}$ by $\bar{\alpha}$. Then, as we show in the proof of Lemma 2, a solution to Equation (24) exists for all $\alpha \leq \bar{\alpha}$.

For $q = \mu$, $\alpha \in [1, \hat{\alpha})$ and $\alpha > \bar{\alpha}$, a price ratio that equates households' expected real returns from both productive sectors does not exist. The reason is the following: With $R_L > R_D$, which follows from $\alpha \in [1, \hat{\alpha})$, a looser leverage constraint α implies that banks' loan supply is higher. BS firms want to use the loans they take to acquire capital from households. Therefore, they need to offer a relative price p_I/p_C for capital that convinces households to provide it to them, instead of acquiring bonds from FS firms. However, Equation (23) shows that a higher p_I/p_C decreases $\hat{\boldsymbol{b}}$ and hence also decreases μ. Therefore, a higher p_I/p_C also has an indirect negative effect on households' expected real return from BS investment, additionally to the direct positive effect. For $\alpha > \bar{\alpha}$, the indirect

[17] For $\hat{\boldsymbol{b}} \geq \overline{\boldsymbol{b}}$ and thus $q = \pi$, we end up in Special case (I) from Subsection 4.3. In fact, one can show that the condition $\hat{\boldsymbol{b}} = \overline{\boldsymbol{b}}$ is equivalent to Condition (19). For $\hat{\boldsymbol{b}} \leq \underline{\boldsymbol{b}}$ and thus $q = \pi - \Delta$, we end up in Special case (II). The condition $\hat{\boldsymbol{b}} = \underline{\boldsymbol{b}}$ is equivalent to Condition (A.24).

negative effect is so strong that no p_I/p_C exists for which households sell the desired amount of capital goods to firms in the BS.[18]

Uniqueness. If a solution to Equation (24) exists, it is typically not unique. From $B > 0, A > 0, \boldsymbol{R_F} > 0$, we can infer that any solution must be positive. In what follows, we assume that a solution exists and focus on the one where $\mu \frac{p_I}{p_C} R_{CB}$ crosses $\boldsymbol{R_F}$ from below:

$$\frac{p_I}{p_C} = \frac{1}{2A} \left(B - \sqrt{B^2 - 4A\boldsymbol{R_F}} \right). \tag{27}$$

The other solution in Equation (26) would imply unintuitive comparative static properties which are in contrast to what we find in Corollary 1. Furthermore, for $\alpha \to 1^{(+)}$, this solution would be incompatible with $\underline{\boldsymbol{b}} < \hat{\boldsymbol{b}} < \overline{\boldsymbol{b}}$ and thus with $q = \mu$, as we show in the proof of Lemma 2.

Lemma 2 summarizes our findings on the existence and uniqueness of an equilibrium price ratio p_I/p_C that equates the real returns from investing in bonds and bank deposits. The proof is in Appendix A. Lemma 3 in the next section shows how $\hat{\alpha}$ and $\bar{\alpha}$ relate to each other.

Lemma 2 (Existence and uniqueness of the equilibrium price ratio)
Let $q = \mu$ and $\alpha \in [1, \hat{\alpha})$. Then:

(i) An equilibrium price ratio p_I/p_C exists only for leverage constraints $\alpha \leq \bar{\alpha}$, where $\bar{\alpha}$ denotes the solution to $B^2 = 4A\boldsymbol{R_F}$, with A and B as defined in Equation (24).

(ii) For $\alpha \leq \bar{\alpha}$, the unique admissible equilibrium price ratio p_I/p_C is given by Equation (27).

Substituting Equation (27) into Equation (23) gives $\hat{\boldsymbol{b}}$ as an expression of exogenous variables only. In line with intuition, Corollary 1 states that a looser leverage constraint α implies a smaller equilibrium portion $\hat{\boldsymbol{b}}$ of monitoring banks and thus also a lower average success probability μ for BS investment. Last, the corollary tells that the equilibrium BS real price of capital p_I/p_C increases in α.

[18] The only way for Equation (14) to hold then is that the price ratio rises beyond the value for which q is capped at $\pi - \Delta$ (i.e., p_I/p_C is such that $\hat{\boldsymbol{b}} \leq \underline{\boldsymbol{b}}$), to $p_I/p_C = \boldsymbol{R_F}/[(\pi - \Delta)R_{CB}]$.

This reflects the fact that a looser leverage constraint does not only lead to higher *nominal* bank-lending volumes, but also affects the amount of *real* resources being provided to the bank-dependent sector.

Corollary 1 (Comparative statics w.r.t. the leverage constraint) *Let $\alpha \in [1, \hat{\alpha})$. The equilibrium effects of a change in the regulatory leverage constraint are given by:*

$$\partial\hat{\boldsymbol{b}}/\partial\alpha < 0, \quad \partial\mu/\partial\alpha < 0, \quad \partial(p_I/p_C)/\partial\alpha > 0.$$

The proof is in Appendix A.

We can now establish Lemma 3, which, for $q = \mu$, determines $\hat{\alpha}$. Denote by α_μ the value of α that solves $\mu\boldsymbol{R_B} = \boldsymbol{R_F}$ (if a solution exists).

Lemma 3 (Threshold value for binding leverage constraints) *Let $q = \mu$. There is an interval $[1, \hat{\alpha})$, for which any $\alpha \in [1, \hat{\alpha})$ implies $R_L > R_D$. The threshold value $\hat{\alpha}$ is given by*

$$\hat{\alpha} = \begin{cases} \alpha_\mu & \textit{if } \alpha_\mu \textit{ exists,} \\ \bar{\alpha} & \textit{otherwise.} \end{cases}$$

Proof. The proof is straightforward. For $q = \mu$, Equations (14) and (17) imply that $R_L > R_D$ is equivalent to $\mu\boldsymbol{R_B} > \boldsymbol{R_F}$. Assumption 2 ensures that $\mu\boldsymbol{R_B} > \boldsymbol{R_F}$, for $\alpha = 1$. As long as $\mu\boldsymbol{R_B} > \boldsymbol{R_F}$, $\mu\boldsymbol{R_B}$ is decreasing in α (cf. Corollary 1), while $\boldsymbol{R_F}$ is increasing in α and approaches infinity for $\alpha \to 1/\boldsymbol{E}$ (cf. Equation (22)). Hence, the Intermediate Value Theorem ensures that there is at most one value of α for which $\mu\boldsymbol{R_B} = \boldsymbol{R_F}$—we denoted this value by α_μ—and if α_μ exists, it has to be between 1 and $\bar{\alpha}$ ($< 1/\boldsymbol{E}$). For all values of α below α_μ, it holds that $\mu\boldsymbol{R_B} > \boldsymbol{R_F}$. If α_μ does not exist, it immediately follows that $\mu\boldsymbol{R_B} > \boldsymbol{R_F}$ for all $\alpha \leq \bar{\alpha}$.

5 Optimal Leverage Constraint in the MC Economy

Until now, we have taken the regulatory leverage constraint α as given. In this section, we find the optimal regulatory leverage constraint, i.e., the value of α the regulator should choose in order to maximize expected aggregate output. For binding regulatory leverage constraints $\alpha \in [1, \hat{\alpha})$, expected aggregate output is given by

$$Y = q\boldsymbol{R_B}\underbrace{\alpha\boldsymbol{E}}_{\boldsymbol{K_B}} + g(\underbrace{1 - \alpha\boldsymbol{E}}_{\boldsymbol{K_F}}). \tag{28}$$

As we will see, the optimal leverage constraint typically implies accepting that not all banks monitor (which was the case in the LF economy), in exchange for a more efficient allocation of capital. Substituting the equilibrium price ratio p_I/p_C, given by Equation (A.12), into Equation (23), yields that $\hat{\boldsymbol{b}}$ is independent of R_{CB} and thus, from Expression (13), also q is independent of R_{CB}. From Equation (28), we can then infer that Y is independent of the CB policy rate R_{CB}. Hence, R_{CB} *does* affect the price ratio and the nominal rates of return, but it *does not* affect the real sphere of the economy.

Maximizing Y with respect to α requires the following first-order condition (FOC):

$$q\boldsymbol{R_B}\frac{\partial \boldsymbol{K_B}}{\partial \alpha} + \frac{\partial q}{\partial \alpha}\boldsymbol{R_B}\boldsymbol{K_B} = g'(\boldsymbol{K_F})\boldsymbol{E}.$$

Making use of Equations (21) and with $g'(\boldsymbol{K_F}) = \boldsymbol{R_F}$, the FOC comes down to

$$q\boldsymbol{R_B} + \boldsymbol{R_B}\alpha\frac{\partial q}{\partial \alpha} = \boldsymbol{R_F}. \tag{29}$$

Denote output Y by Y_π, for $q = \pi$, and by Y_Δ, for $q = \pi - \Delta$. Then, the FOC simplifies to $\pi\boldsymbol{R_B} = \boldsymbol{R_F}$ in the former and $(\pi - \Delta)\boldsymbol{R_B} = \boldsymbol{R_F}$ in the latter case. Furthermore, denote output Y by Y_μ, for $q = \mu$, where μ is given by Expression (13), $\hat{\boldsymbol{b}}$ is given by Equation (23) and p_I/p_C is given by Equation (27). The following subsection focuses on this last case.[19]

[19] Which case actually applies depends on whether $\hat{\boldsymbol{b}} \geq \overline{\boldsymbol{b}}$ ($\Rightarrow Y = Y_\pi$), $\hat{\boldsymbol{b}} \leq \underline{\boldsymbol{b}}$ ($\Rightarrow Y = Y_\Delta$) or $\underline{\boldsymbol{b}} < \hat{\boldsymbol{b}} < \overline{\boldsymbol{b}}$ ($\Rightarrow Y = Y_\mu$). As $\hat{\boldsymbol{b}}$ depends on α itself, these conditions depend on α as well. In Subsection 5.2, we provide a set of conditions that ensures that the α that maximizes Y_μ indeed satisfies $\underline{\boldsymbol{b}} < \hat{\boldsymbol{b}} < \overline{\boldsymbol{b}}$.

5.1 Locally optimal leverage constraint

Denote by α^* the α that solves the FOC (29) with q given by μ. To verify whether α^* indeed (uniquely) maximizes Y_μ, Proposition 4 characterizes Y_μ as a function of α.

Proposition 4 (Locally optimal leverage constraint)
Let $g'''(\boldsymbol{K_F}) > 0$. Then,

(i) Y_μ is strictly concave in $\alpha \in [1, \bar{\alpha})$, i.e., there is at most one $\alpha \in [1, \bar{\alpha})$ that solves the FOC (29), and if it exists, it constitutes a maximum.

(ii) There is an $\alpha \in [1, \bar{\alpha})$ that solves the FOC (29), iff

$$g'(1-E) \leq \frac{\mu_1^2 \boldsymbol{R_B}}{\mu_1 + \frac{\Delta^2 \boldsymbol{R_B}}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}}. \tag{30}$$

The proof is given in Appendix A. It also shows that the condition $g'''(\boldsymbol{K_F}) > 0$ is sufficient, but not necessary. This condition is met, e.g., by a standard production function of the form $g(\boldsymbol{K}_F) = \boldsymbol{K_F}^\beta$, with $0 < \beta < 1$. Condition (30) ensures that an increase in α, starting from $\alpha = 1$, increases Y_μ. A necessary condition for this is $g'(1 - \boldsymbol{E}) < \mu_1 \boldsymbol{R_B}$ (cf. Assumption 2), but it is not sufficient, since one has to take into account that any increase in α negatively affects the average success probability μ. If $g'''(\boldsymbol{K_F}) > 0$ and Condition (30) holds, α^* (> 1) uniquely maximizes Y_μ. We then call α^* the (locally) optimal leverage constraint.

Remember that we restricted attention to leverage constraints that are binding for all banks b, i.e., to values of α that imply $R_L > R_D$ and thus belong to the interval $\alpha \in [1, \hat{\alpha})$. Corollary 2 establishes $\alpha^* \in [1, \hat{\alpha})$.

Corollary 2 (Binding locally optimal leverage constraint)
Let the conditions of Proposition 4 hold. The leverage constraint α^* is binding for all banks b, i.e., $\alpha^* \in [1, \hat{\alpha})$.

Proof. From Lemma 3, either $\hat{\alpha} = \alpha_\mu$, if α_μ exists, or $\hat{\alpha} = \bar{\alpha}$, if α_μ does not exist. From Proposition 4*(ii)* and the accompanying proof, it follows that $\alpha^* \in [1, \bar{\alpha})$. Hence, if $\hat{\alpha} = \bar{\alpha}$, the corollary holds. Consider now the case of $\hat{\alpha} = \alpha_\mu$. From

FOC (29), it holds that, at $\alpha = \alpha^*$,

$$\boldsymbol{R_F} = \mu \boldsymbol{R_B} \left(1 + \underbrace{\frac{\alpha^*}{\mu} \frac{\partial \mu}{\partial \alpha}}_{\varepsilon_{\mu,\alpha^*}} \right), \tag{31}$$

where $\varepsilon_{\mu,\alpha^*}$ is the elasticity of μ with respect to α, evaluated at $\alpha = \alpha^*$. As long as $\mu \boldsymbol{R_B} > \boldsymbol{R_F}$, from Corollary 1 and Equation (22), it follows that $\boldsymbol{R_F}$ is increasing in α, $\mu \boldsymbol{R_B}$ is decreasing in α and $\varepsilon_{\mu,\alpha^*} < 0$. Then, by the Intermediate Value Theorem, α^*, as given by Equation (31), is smaller than the value of α that equates $\boldsymbol{R_F}$ and $\mu \boldsymbol{R_B}$, i.e., it holds that $\alpha^* < \alpha_\mu$.[20]

Corollary 2 implies that when choosing the optimal leverage constraint α^*, the regulator accepts that capital is not allocated perfectly efficiently, i.e., that the expected marginal product of capital in the BS exceeds that of the FS, in exchange for higher monitoring activity.

5.2 Globally optimal leverage constraint

The last subsection characterized α^* as the α that maximizes Y_μ. In this subsection, we explore a set of conditions that establishes α^* as the globally optimal leverage constraint.

To do this, we require some additional notation. Denote the α that solves $\hat{\boldsymbol{b}} = \overline{\boldsymbol{b}}$ by α_m and the α that solves $\hat{\boldsymbol{b}} = \underline{\boldsymbol{b}}$ by α_s, where $\hat{\boldsymbol{b}}$ is given by Equation (23) with p_I/p_C according to Equation (27). Since $\hat{\boldsymbol{b}}$ is strictly decreasing in α (cf. Corollary 1), it follows that for $\alpha \leq \alpha_m$, all banks monitor, for $\alpha \geq \alpha_s$, all banks shirk and for $\alpha_m < \alpha < \alpha_s$, some banks monitor while some others shirk. Analogously, denote the α that solves $(\pi - \Delta)\boldsymbol{R_B} = \boldsymbol{R_F}$, which, for $q = \pi - \Delta$, is equivalent to $R_L = R_D$, by α_Δ.

Global output function. Expected aggregate output as a function of the regulatory leverage constraint α, with $\alpha \in [1, \infty)$, is denoted by Y^g (henceforth called the "global output function"). For all values of α that imply $R_L \geq R_D$ if banks leverage up as much as possible, Y^g is simply given by Y according to Equation

[20] Note that while, for $\alpha \geq \hat{\alpha} \, (= \alpha_\mu)$ or, equivalently, for $\mu \boldsymbol{R_B} \leq \boldsymbol{R_F}$, Corollary 1 is silent on the sign of $\partial \mu / \partial \alpha$, $\partial \mu / \partial \alpha$ does not converge to zero for $\alpha \to \alpha_\mu^{(-)}$. In fact, $\lim_{\alpha \to \alpha_\mu^{(-)}} \partial \mu / \partial \alpha$ is strictly smaller than zero.

(28). For all α's that would imply $R_L < R_D$ if banks leveraged up as much as possible, Y^g cannot be determined exactly but is certainly non-increasing (indicated by "$\searrow$"). It is non-increasing because any leverage constraint beyond the one for which R_L equals R_D would only induce an overallocation of capital from non-monitoring banks to the BS.[21]

An explicit expression of Y^g requires tedious case distinctions. The reason is the following. Since p_I/p_C, according to Equation (27), exists only for $\alpha \leq \bar{\alpha}$, existence of α_s and α_μ is not guaranteed. Furthermore, the pattern of Y^g differs, depending on whether $\alpha_s \gtreqless \alpha_\mu$ and whether $\alpha_\Delta \gtreqless \alpha_s$. In Gersbach and Zelzner (2022), we perform these case distinctions in detail. To illustrate one possible case, assume that both α_s, α_μ exist, and that $\alpha_s < \alpha_\mu$ and $\alpha_\Delta \geq \alpha_s$. Under these assumptions, we can simply state Y^g as

$$Y^g = \begin{cases} Y_\pi & \text{for } 1 \leq \alpha \leq \alpha_m, \\ Y_\mu & \text{for } \alpha_m < \alpha < \alpha_s, \\ Y_\Delta & \text{for } \alpha_s \leq \alpha \leq \alpha_\Delta, \\ \searrow & \text{for } \alpha > \alpha_\Delta. \end{cases}$$

In fact, we can derive a necessary and sufficient set of conditions that establishes α^* as the unique leverage constraint that maximizes Y^g in all possible cases. This result is summarized in the following proposition.

Proposition 5 (Globally optimal leverage constraint)
The leverage constraint α^* implies $\pi - \Delta < \mu < \pi$ and maximizes Y^g, if and only if

$$(i)\ \left.\frac{\partial Y_\mu}{\partial \alpha}\right|_{\alpha=\alpha_m} > 0, \quad (ii)\ Y_\mu(\alpha^*) > Y_\Delta(\alpha_\Delta).$$

The proof is in Appendix A. Conditions *(i)* and *(ii)* ensure that the globally optimal leverage constraint implies positive fractions of non-monitoring and monitoring banks, respectively.[22]

[21] Since non-monitoring banks enjoy private benefits that scale with lending, they may want to continue to lend as much as possible also for $R_L < R_D$. But even they eventually lose their incentive to lend, as the spread $R_D - R_L$ becomes large. Furthermore, the constraint that banks have to be able to repay the CB applies (cf. Assumption 3).

[22] We note that Condition *(ii)* implies $(\partial Y_\mu/\partial\alpha)|_{\alpha=\alpha_s} < 0$ (if α_s exists).

We can further evaluate Condition *(i)* of the proposition. By definition, α_m is determined by $\hat{\boldsymbol{b}} = \overline{\boldsymbol{b}}$ and $\alpha = \alpha_m$ implies $q = \pi$. Making use of Equation (23) and $p_I/p_C = \boldsymbol{R_F}/(\pi R_{CB})$ yields that α_m is implicitly given by

$$g'(1-\alpha_m \boldsymbol{E})\left(1-\frac{1}{\alpha_m}\right) = \frac{\pi}{\Delta}(\Delta \boldsymbol{R_B} - \overline{\boldsymbol{b}}).$$

From Equation (29), $(\partial Y_\mu/\partial\alpha)|_{\alpha=\alpha_m} > 0$ and thereby Condition *(i)* holds, iff

$$\pi \boldsymbol{R_B} - g'(1-\alpha_m \boldsymbol{E}) > -\boldsymbol{R_B}\alpha_m \left.\frac{\partial\mu}{\partial\alpha}\right|_{\alpha=\alpha_m},$$

i.e., if, at $\alpha = \alpha_m$, the benefits of a more efficient allocation of capital resulting from a marginal increase in α (as given by the l.h.s.) outweigh the associated drawback from lower monitoring activity (as given by the r.h.s.).

We next provide a set of normalizing assumptions which allows us to assess the conditions stated in Proposition 5 further, without having to refer to any case distinctions.

Corollary 3 (Normalization)
Let the following conditions jointly hold:

$$\overline{\boldsymbol{b}} = \Delta\boldsymbol{R_B}, \quad \underline{\boldsymbol{b}} = \Delta\boldsymbol{R_B}\boldsymbol{E}/[1-(g')^{-1}\{(\pi-\Delta)\boldsymbol{R_B}\}], \quad \Delta \leq 0.5\pi.^{23} \tag{32}$$

Then,

(i) α_s *and* α_μ *exist and it holds that* $\alpha_m = 1$ *and* $\alpha_s = \alpha_\Delta = \alpha_\mu$*,*

(ii) $\alpha^* \in (\alpha_m, \alpha_s)$ *maximizes* Y^g*, if, additionally,*

$$g'(1-\boldsymbol{E}) < \frac{\pi^2 \boldsymbol{R_B}}{\pi + \frac{\Delta\boldsymbol{R_B}}{\Delta\boldsymbol{R_B}-\underline{\boldsymbol{b}}}\Delta}. \tag{33}$$

The proof is given in Appendix A. The first two conditions in the set of conditions in (32) normalize the interval $[\underline{\boldsymbol{b}}, \overline{\boldsymbol{b}}]$, such that $\alpha_m = 1$ and $\alpha_s = \alpha_\Delta = \alpha_\mu$. The third condition ensures that α_s and α_μ exist. Part *(ii)* of the corollary tells us that the set of conditions in (32), combined with Condition (33), constitutes a

[23] Note that, from Equation (A.23), $\boldsymbol{E}/[1-(g')^{-1}\{(\pi-\Delta)\boldsymbol{R_B}\}] = 1/\alpha_\Delta$, which, under Assumption 2, is smaller than one. Hence, the set of conditions in (32) satisfies $\underline{\boldsymbol{b}} < \overline{\boldsymbol{b}}$.

simple set of *sufficient* closed-form conditions solely in exogenous variables that implies that Conditions *(i)–(ii)* in Proposition 5 are satisfied.

5.3 Comparing the MC economy to the LF economy

We can now compare aggregate outcomes in the MC economy to those in the LF economy. Proposition 6 states our main result.

Proposition 6 (Why bank money creation?)
Let the conditions given in Proposition 5 hold. Then, the MC economy with a regulatory leverage constraint α^* features greater expected aggregate output than the LF economy.

Corollary 3 gives a simple set of sufficient conditions for the conditions stated in Proposition 5. The proof of Proposition 6 follows from our previous results. According to Proposition 3, the MC economy and the all-monitor LF economy do equally well for a leverage constraint $\alpha = \alpha_m$. Hence, if the conditions of Proposition 5 hold, output in the MC economy with a leverage constraint α^* (with $\alpha_m < \alpha^* < \alpha_s$) clearly exceeds output in the LF economy.[24]

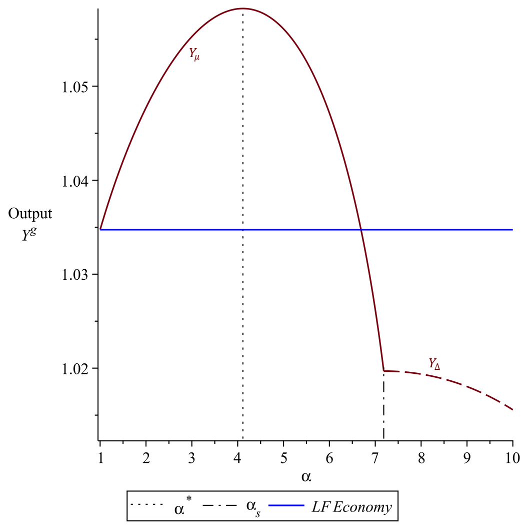
Figure 2 illustrates the result. Parameters are chosen such that the set of conditions in Corollary 3 holds. The output function Y^g is smooth for $\alpha \in [1, \alpha_s)$ and shows a kink at $\alpha = \alpha_s$, as q is capped at $\pi - \Delta$ beyond this point. Output for $\alpha > \alpha_s (= \alpha_\mu = \alpha_\Delta)$ is dashed since we certainly know that it must be non-increasing beyond this point but cannot determine how exactly it behaves. Obviously, α^* maximizes Y^g. The "LF Economy" line shows output in the all-monitor equilibrium of the LF economy. Note that the introduction of a regulatory leverage constraint in the LF economy would not offer any scope for improvement, since households restrict bank deposits such that all banks monitor anyway.[25]

Although output in the MC economy exceeds output in the LF economy, it obviously falls short of the first-best. In Gersbach and Zelzner (2022), we also

[24] As explained in Footnote 11, we cannot strictly rule out an all-shirk equilibrium in the LF economy. Such an all-shirk LF economy, however, would certainly also fare worse than the MC economy: while an all-shirk equilibrium of the LF economy could easily be replicated in the MC economy by setting the appropriate α with $\alpha \geq \alpha_s$, under the conditions of Proposition 5 it is optimal not to do so but instead to set $\alpha = \alpha^*$, with $\alpha^* < \alpha_s$.

[25] At most, such a leverage constraint could strictly rule out an all-shirk equilibrium and thus establish the all-monitor equilibrium as the unique equilibrium of the LF economy.

Figure 2: Output Y^g as a function of the regulatory leverage constraint α



consider a second-best LF economy with symmetric information, where households can stick to funding constraints tailor-made for individual banks. Output in the MC economy typically also falls short of output in such a second-best LF economy. Therefore, while the MC economy can alleviate the problem of asymmetric information on bank characteristics by striking an optimal balance on the trade-off between high aggregate monitoring and an efficient allocation of capital, it typically does not solve the problem altogether.

Unobservable bank heterogeneity. The reason why the MC economy with an appropriately set leverage constraint is superior to the LF economy in our model is unobservable heterogeneity among banks, combined with agents' price-taking behavior.[26]

In essence, unobservable bank-heterogeneity is the reason why the differing sequence of events in the LF economy, compared to the MC economy, is important. In the LF economy, the first step in banking intermediation is that households provide funding to banks. Because of asymmetric information, households cannot distinguish between banks and thus, according to Condition (2), they strongly restrict the amount of capital they provide to all of them to ensure that the banks

[26]If all banks were the same, also the MC economy could only either achieve an all-monitor or an all-shirk equilibrium. If banks were heterogeneous, but their types observable, households in the LF economy could incentivize banks to monitor on a bank-by-bank basis. The MC economy could achieve exactly the same by setting appropriate bank-individual leverage constraints, which typically should also be optimal.

have the incentives to monitor. In contrast, the first step in the MC economy is that banks can initiate lending on their own, being restricted only by the regulatory leverage constraint. Price-taking households do not take into account how the amount of capital goods they sell to BS firms affects banks' incentive constraint (12) for monitoring.

Finally, we note that the regulator sets the optimal leverage constraint in the MC economy by taking into account how a change of this constraint will affect the banks' monitoring decisions, capital prices and the allocation of capital in the economy. In particular, the regulator not only takes into account how channeling more funds to the BS may decrease monitoring incentives, but also how it may lead to a more efficient allocation of capital in the economy. In the LF economy, price-taking households can only take into account how changing their investment may affect the banks' monitoring incentives, but they are, of course, not concerned about capital allocation in the economy and capital price changes.

On the fragility of bank money creation's benefits. As we have shown, output in the MC economy with $\alpha = \alpha^*$ exceeds output in the LF economy. However, while equilibrium in the LF economy is formed only through market mechanisms, the MC economy requires a well informed regulator who is willing and able to enforce a leverage constraint α^*. This renders the optimal outcome in the MC economy fragile. In Figure 2, we can see that while the MC economy at $\alpha = \alpha^*$ is superior to the LF economy, this is not true for all α. Thus, the regulator needs precise knowledge about all relevant economic fundamentals to accurately set the optimal leverage constraint. If that is not the case or if the regulator sets the leverage constraint wrong for other reasons, output in the MC economy may as well fall short of output in the LF economy.

6 Bank-risk Assessment and Risk-sensitive Capital Requirements

In this section, we show that it is possible to further improve outcomes in the MC economy by combining bank-risk assessments with risk-sensitive capital requirements. Assume that the regulator delegates these assessments to the CB, which performs them at the end of period $t = 1$ and immediately informs the regulator of the results.

6.1 Perfect risk assessment

As the benchmark case, assume that bank-risk assessment perfectly reveals the default probability of each bank's loans: $1 - \pi$ or $1 - (\pi - \Delta)$. Observing these probabilities, the regulator can perfectly infer whether the bank monitored or not. For simplicity, assume that risk assessment is costless. Then, at the beginning of period $t = 1$, the regulator can announce that:

(i) high-risk banks, i.e., banks with default probability $1 - (\pi - \Delta)$, have to comply with a maximum leverage ratio of $\alpha = \alpha_m$, and

(ii) low-risk banks, i.e., banks with default probability $1 - \pi$, have to comply with a maximum leverage ratio of $\alpha = \alpha_\pi$, where α_π is implicitly given by $\pi \boldsymbol{R_B} = \boldsymbol{R_F}$, which yields

$$\alpha_\pi = \frac{1 - (g')^{-1}(\pi \boldsymbol{R_B})}{\boldsymbol{E}}.$$

At the same time, the regulator also announces that the CB will assess all banks' risk types and control their compliance, with the respective leverage constraints at the end of period $t = 1$. In case of non-compliance, the regulatory authority announces that it will levy a large penalty, which banks want to avoid by all means (in the medium- to long-run, these banks would also be forced to de-leverage). We assume that the regulatory authority can commit to its statements and, hence, this announcement is credible.

With (i), all banks decide to monitor. They know that if they don't, they will have to comply with a leverage constraint for which they would have been better off monitoring in the first place. With all banks monitoring, (ii) sets the leverage constraint such that capital is allocated efficiently. Hence, we can state the following proposition.

Proposition 7 (Bank-risk assessment and capital requirements)
Suppose that the CB can perfectly assess the default risk of the banks' credit portfolios. Then, first-best is achieved in the MC economy by introducing risk-sensitive

leverage constraints α_m and α_π for high-risk and low-risk banks, respectively.

Interestingly, we achieve first-best despite the fact that the regulator cannot observe individual bank types. The threat of a sufficiently strict leverage constraint for high-risk (i.e., non-monitoring) banks already ensures that all banks have an incentive to monitor. The reason why this results in a first-best scenario is that the strict leverage constraint $\alpha = \alpha_m$ only serves as a threat but never actually has to be put into effect.

6.2 Imperfect risk assessment

In the benchmark case above, we assumed that the CB can (a) perfectly assess a bank's credit risks, and then (b) perfectly infer whether the bank monitored or not. Of course reality shows a noisier picture. If the effect of monitoring efforts on credit risks is not entirely deterministic, (b) is no longer possible. Naturally, we would still expect a correlation, but inference would be imperfect.

Even worse, if risk assessment is imperfect and hence (a) is violated, banks could have an incentive to first skip monitoring efforts and then try to deceive the CB into believing that they are in fact low-risk. Balance sheet window-dressing, for instance, is both within the law and common among banks (see, e.g., Allen and Saunders, 1992; Shaffer and Yang, 2010). Disguising credit risk by re-packaging loans into structured products was common bank practice in the run-up of the financial crisis of 2007-08. All of this can generate a situation where banks are often misclassified as low-risk, while in fact they are not. The problem with risk-sensitive capital requirements as stated above then would be that banks that are perceived as low-risk face only a leverage constraint α_π. As α_π is obviously less strict than the optimal leverage constraint α^* without bank-risk assessments, this implies that risk-sensitive capital requirements could backfire and in fact worsen outcomes in an MC economy if banks can easily deceive risk assessments.

Underestimated bank credit risks were considered one of the main causes for the financial crisis of 2007-08. Many new regulations have been put in place since then. For instance, Basel III introduced generally stricter capital requirements. Furthermore, regulators and central banks have invested in better capabilities in bank-risk assessment and stress-testing. We can thus assume that while bank-risk assessments certainly are still imperfect, they now tend to be less prone to failure. Taking this into account, bank-risk assessments, combined with risk-sensitive capital requirements might not result in first-best as in the benchmark case, but could still contribute to further improvements in an MC economy.

7 Conclusion

We develop a model to illustrate the merits of a monetary system with bank money creation over an economy with banks as simple intermediaries of loanable funds. In the presence of bank heterogeneity and potential bank-level moral hazard, the fact that banks do not need household funding to initiate lending leads to higher lending volumes, a more efficient allocation of capital and, under a suitable regulatory leverage constraint, to higher economic output overall. Bank-risk assessments, combined with risk-sensitive capital requirements, can improve outcomes further and, under certain conditions, even achieve the first-best allocation.

Policy-wise, we provide a rationale for bank money creation and thus offer an argument against proposals to abolish this privilege for banks. In this regard, our findings also matter for the ongoing discussion on the introduction of CBDCs. In particular, central banks should be careful in which precise manner such a digital currency would be implemented, so that the benefits of private money creation in our current two-tier monetary system are not lost. With regard to economic modeling, the differing outcomes in the MC and LF economies suggest that the standard LF approach to banking should not be considered a simple short-cut to the MC approach in settings with heterogeneous banks and financial frictions at bank level.

A Proofs

Proof of Proposition 2. Let $R_F - \pi R_B + \pi \bar{b}/\Delta > 0$, otherwise the constraint in the maximization problem in (4) never binds. Then, rewriting this constraint yields

$$k \leq \frac{eR_F}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}}. \tag{A.1}$$

Banks' objective function in the maximization problem in (4) is linear in k. This implies that banks lend as much as possible if $\pi R_B > R_F$, they do not lend at all if $\pi R_B < R_F$ and they lend an arbitrary amount if $\pi R_B = R_F$. As we focus on equilibria where positive amounts of capital are provided to the bank-dependent sector, let $\pi R_B \geq R_F$. We distinguish two cases: *(i)* $R_F = \pi R_B$, and *(ii)* $R_F < \pi R_B$.

Case *(i)*. Assume that $R_F = \pi R_B$. Then, Condition (A.1) is given by

$$k \leq \frac{\Delta e R_B}{\bar{b}}.$$

For the aggregate economy, this implies

$$K_B \leq \frac{\Delta e R_B}{\bar{b}}(\bar{b} - \underline{b}).$$

Hence, an equilibrium with $R_F = \pi R_B$ is consistent with Condition (A.1) if and only if

$$e \geq \frac{K_B \bar{b}}{\Delta R_B(\bar{b} - \underline{b})}.$$

Since $K_B = 1 - K_F$, $K_F = (g')^{-1}(R_F)$ and $E = (\bar{b} - \underline{b})e$, we can restate this condition in exogenous variables only:

$$E \geq \frac{\bar{b}\,(1 - (g')^{-1}(\pi R_B))}{\Delta R_B}. \tag{A.2}$$

Since $g'(\cdot)$ is strictly monotonically decreasing, the inverse $(g')^{-1}(\cdot)$ exists. If Condition (A.2) holds, the incentive constraint (A.1) is non-binding at $R_F = \pi R_B$. Hence, in this case, $R_F = \pi R_B$ constitutes the equilibrium value of R_F.

Case *(ii)*. Assume that $R_F < \pi R_B$. Then, as much capital as possible flows into the BS, i.e., Condition (A.1) is binding:

$$k = \frac{eR_F}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}}. \tag{A.3}$$

In aggregate, this implies

$$K_B = \frac{eR_F}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}}(\bar{b} - \underline{b}). \tag{A.4}$$

Using $K_B = 1 - K_F$ and $g'(K_F) = R_F$, this yields

$$R_F = g' \left\{ 1 - \frac{eR_F(\bar{b} - \underline{b})}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}} \right\}. \tag{A.5}$$

As the l.h.s. is linearly increasing in R_F, and one can show that the r.h.s. is strictly monotonically decreasing in R_F, there is a unique value of R_F that solves this equation (see the Auxiliary Lemma A.1 below). For $\pi R_B > R_F$, the l.h.s. evaluated at $R_F = \pi R_B$ has to be greater than the r.h.s. evaluated at $R_F = \pi R_B$:

$$\pi R_B > g' \left\{ 1 - \frac{\Delta e R_B(\bar{b} - \underline{b})}{\bar{b}} \right\}.$$

Solving for e and using $E = (\bar{b} - \underline{b})e$ yields Condition (5). If this condition holds, there is a value of R_F ($< \pi R_B$) that solves Equation (A.5) and hence constitutes the equilibrium value of R_F.

Auxiliary Lemma A.1

There is a unique value of R_F that solves Equation (A.5) and satisfies $R_F > \pi R_B - (\pi \bar{b}/\Delta)$.

Proof. We prove Auxiliary Lemma A.1 by using the Intermediate Value Theorem. The l.h.s. of Equation (A.5) is linearly increasing in R_F. Next, we show that the r.h.s. of Equation (A.5) is decreasing in R_F. Let

$$f(R_F) \equiv \frac{ER_F}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}}.$$

Then, since $g''(\cdot) < 0$, the r.h.s. of Equation (A.5) is decreasing in R_F if and only if $f(R_F)$ is decreasing in R_F. This is the case, iff

$$\frac{E}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}} \left[1 - \frac{R_F}{R_F - \pi R_B + \frac{\pi \bar{b}}{\Delta}} \right] \leq 0. \tag{A.6}$$

As assumed at the beginning of the proof to Proposition 2, $R_F - \pi R_B + \pi \bar{b}/\Delta > 0$. Assumption 1 implies that $-\pi R_B + \pi \bar{b}/\Delta \leq 0$ and thus, from Equation (A.3), ensures that $k \geq e$ and, in aggregate, $K_B \geq E$. Then, from Equation (A.4) we also obtain $R_F/(R_F - \pi R_B + \pi \bar{b}/\Delta) \geq 1$. It follows that Inequality A.6 holds.

Finally, note that the r.h.s. of Equation (A.5) goes to infinity for R_F to $[\pi R_B - (\pi \bar{b}/\Delta)]/(1-E)\,(> \pi R_B - (\pi \bar{b}/\Delta))$ from above. Therefore, an $R_F\,(> \pi R_B - (\pi \bar{b}/\Delta))$ that solves Equation (A.5) exists. ■

Proof of Lemma 2. From $\alpha < \hat{\alpha}$ follows that Expression (21) applies and thus also Equations (22)–(26) apply.

Part *(i)*. Whether a solution to Equation (24) exists, depends on whether $B^2 - 4A\boldsymbol{R_F} \geq 0$. This is the case, iff

$$\begin{aligned} R_{CB}^2 \mu_1^2 - 4 R_{CB}^2 \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \left(1 - \frac{1}{\alpha}\right) \boldsymbol{R_F} \geq 0 \\ \mu_1^2 \geq 4 \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} g'(1 - \alpha \boldsymbol{E}) \left(1 - \frac{1}{\alpha}\right). \end{aligned} \tag{A.7}$$

For $\alpha \to 1^{(+)}$, the r.h.s. goes to zero and Condition (A.7) obviously holds. For $\alpha \to \boldsymbol{1/E}$, the r.h.s. goes to infinity (as $g'(0) = \infty$) and the inequality does not hold. As the l.h.s. is independent of α and the r.h.s. is strictly increasing in α, by the Intermediate Value Theorem there is exactly one value of α in $(1, 1/\boldsymbol{E})$ for which Condition (A.7) holds with equality. If we denote this value by $\alpha = \bar{\alpha}$, a solution exists for all $\alpha \leq \bar{\alpha}$ and it does not exist for all $\alpha > \bar{\alpha}$.

Part *(ii)*. Substituting the second solution in Equation (26), i.e.,

$$\frac{p_I}{p_C} = \frac{1}{2A} \left(B + \sqrt{B^2 - 4A\boldsymbol{R_F}} \right), \tag{A.8}$$

into $\hat{\boldsymbol{b}}$, yields

$$\hat{\boldsymbol{b}} = \Delta \boldsymbol{R_B} - \frac{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}{2\Delta} \left(\mu_1 + \sqrt{\mu_1^2 - \frac{4A\boldsymbol{R_F}}{R_{CB}^2}} \right). \tag{A.9}$$

For α approaching one from above, $A \to 0^{(+)}$. Using this and substituting μ_1 as defined in Expression (25) into Equation (A.9) yields

$$\lim_{\alpha \to 1^{(+)}} \hat{\boldsymbol{b}} = \overline{\boldsymbol{b}} - \frac{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}{\Delta} \pi.$$

It follows that, for $\alpha \to 1^{(+)}$, $\hat{\boldsymbol{b}} > \underline{\boldsymbol{b}}$ requires $\pi - \Delta < 0$, which can never be the case by assumption. Hence, p_I/p_C as given by Equation (A.8) is not an admissible solution to Equation (26). The only admissible solution to Equation (26), and thus the equilibrium price ratio, is then given by Equation (27).

Note that, besides the argument just made, the solution for p_I/p_C as given by Equation (A.8) would also imply that $\mu(p_I/p_C)R_{CB}$ crosses $\boldsymbol{R_F}$ from above, which in turn would yield unintuitive comparative static properties that are in contrast to what we find in Corollary 1. ■

Proof of Corollary 1. We first show that $\hat{\boldsymbol{b}}$ is decreasing in α. For $q = \pi$ or $q = \pi - \Delta$, the equilibrium price ratio p_I/p_C is obviously increasing in α, since $\partial \boldsymbol{R_F}/\partial\alpha > 0$ by Equation 22. Hence, in these cases, from Equation (23) we immediately see that $\hat{\boldsymbol{b}}$ is decreasing in α. For $q = \mu$, plugging the equilibrium price ratio given by Equation (27) into Equation (23) yields

$$\hat{\boldsymbol{b}} = \Delta \boldsymbol{R_B} - \frac{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}{2\Delta} \left(\mu_1 - \sqrt{\mu_1^2 - \frac{4A\boldsymbol{R_F}}{R_{CB}^2}} \right).$$

As μ_1 is independent of α, and A as well as $\boldsymbol{R_F}$ are increasing in α, we see that $\hat{\boldsymbol{b}}$ is decreasing in α.

With Expression (13), $\partial\mu/\partial\alpha < 0$ follows immediately from $\partial\hat{\boldsymbol{b}}/\partial\alpha < 0$. With regard to the equilibrium price ratio, we already assessed $\partial(p_I/p_C)/\partial\alpha > 0$ for $q = \pi$ or $q = \pi - \Delta$. For $q = \mu$, total differentiation of Equation (24) yields

$$\frac{d\frac{p_I}{p_C}}{d\alpha} = \frac{R_{CB}^2 \left(\frac{p_I}{p_C}\right)^2 \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \frac{1}{\alpha^2} - \boldsymbol{E} g''(1 - \alpha \boldsymbol{E})}{R_{CB} \left(\pi - \Delta \frac{\overline{\boldsymbol{b}} - \Delta \boldsymbol{R_B}}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \right) - 2 \frac{p_I}{p_C} R_{CB}^2 \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \left(1 - \frac{1}{\alpha}\right)}. \tag{A.10}$$

As $g''(\cdot) < 0$, the numerator is always positive. Hence, $\partial(p_I/p_C)/\partial\alpha > 0$, if

$$R_{CB}\mu_1 - 2\frac{p_I}{p_C}R_{CB}^2\frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right) > 0. \tag{A.11}$$

From Equation (27), the equilibrium price ratio p_I/p_C is given by

$$\frac{p_I}{p_C} = \frac{\mu_1 - \sqrt{\mu_1^2 - 4\boldsymbol{R_F}\frac{\Delta^2}{\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right)}}{2R_{CB}\frac{\Delta^2}{\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right)}. \tag{A.12}$$

Then, Inequality (A.11) holds, if

$$\mu_1 R_{CB} - \left(\mu_1 R_{CB} - R_{CB}\sqrt{\mu_1^2 - 4\boldsymbol{R_F}\frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right)}\right) > 0$$
$$\sqrt{\mu_1^2 - 4\boldsymbol{R_F}\frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right)} > 0.$$

This either holds or an equilibrium price ratio consistent with $q = \mu$ does not exist in the first place (cf. Appendix A). Hence, p_I/p_C is increasing in α. ■

Proof of Proposition 4. We first prove Part *(i)* of the Proposition by showing that Y_μ is strictly concave in α. After that, we proceed with a proof for Part *(ii)*.

Part *(i)*. From Equation (28), we obtain

$$\frac{\partial^2 Y_\mu}{\partial\alpha^2} = \boldsymbol{E}\left[\boldsymbol{R_B}\frac{\partial\mu}{\partial\alpha} + \boldsymbol{R_B}\left(\frac{\partial\mu}{\partial\alpha} + \alpha\frac{\partial^2\mu}{\partial\alpha^2}\right) + \boldsymbol{E}g''(\boldsymbol{K_F})\right].$$

This is smaller than zero, if

$$2\boldsymbol{R_B}\frac{\partial\mu}{\partial\alpha} + \alpha\boldsymbol{R_B}\frac{\partial^2\mu}{\partial\alpha^2} + \boldsymbol{E}g''(\boldsymbol{K_F}) < 0. \tag{A.13}$$

We first calculate $\partial\mu/\partial\alpha$. From the definition of μ_1 in Equation (25) and the definition of μ in Expression (13), with $\hat{\boldsymbol{b}}$ given by Equation (23), we obtain

$$\mu = \mu_1 - \frac{\Delta\boldsymbol{R_B}}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\Delta + \frac{\hat{\boldsymbol{b}}}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\Delta,$$
$$\mu = \mu_1 - \frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right)R_{CB}\frac{p_I}{p_C}. \tag{A.14}$$

Substituting p_I/p_C from Equation (A.12) into Equation (A.14) yields

$$\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\underbrace{\sqrt{\mu_1^2 - 4\boldsymbol{R_F}\frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left(1 - \frac{1}{\alpha}\right)}}_{=:\Psi} \tag{A.15}$$

and thereby

$$\Psi = 2\mu - \mu_1. \tag{A.16}$$

By taking the derivative of μ, given by Equation (A.15), with respect to α and making use of Equation (A.16), we obtain

$$\frac{\partial \mu}{\partial \alpha} = \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left[-\frac{\boldsymbol{R_F}}{\alpha^2(2\mu - \mu_1)} + \left(1 - \frac{1}{\alpha}\right)\frac{\boldsymbol{E}g''(\boldsymbol{K_F})}{2\mu - \mu_1}\right]. \tag{A.17}$$

From Expression (A.17), we can also calculate $\partial^2\mu/\partial\alpha^2$:

$$\begin{aligned}\frac{\partial^2 \mu}{\partial \alpha^2} = {}& \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\left\{\frac{\boldsymbol{E}g''(\boldsymbol{K_F})}{\alpha^2(2\mu - \mu_1)} + \frac{\boldsymbol{R_F}}{\alpha^4(2\mu - \mu_1)^2}\left[2\alpha(2\mu - \mu_1) + 2\alpha^2\frac{\partial \mu}{\partial \alpha}\right]\right\} + \\ & + \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\boldsymbol{E}\underbrace{\left\{-\boldsymbol{E}g'''(\boldsymbol{K_F})\frac{1 - \frac{1}{\alpha}}{2\mu - \mu_1} + g''(\boldsymbol{K_F})\left[\frac{1}{\alpha^2(2\mu - \mu_1)} - \frac{2(1 - \frac{1}{\alpha})}{(2\mu - \mu_1)^2}\frac{\partial \mu}{\partial \alpha}\right]\right\}}_{=:\Phi}.\end{aligned} \tag{A.18}$$

Substituting Expression (A.18) into the l.h.s. of Condition (A.13), we obtain

$$\begin{aligned}2\boldsymbol{R_B}\frac{\partial \mu}{\partial \alpha} + \frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\boldsymbol{R_B}\left\{\frac{2\boldsymbol{R_F}}{\alpha^2(2\mu - \mu_1)} + \frac{2\boldsymbol{R_F}}{\alpha(2\mu - \mu_1)^2}\frac{\partial \mu}{\partial \alpha} + \frac{\boldsymbol{E}g''(\boldsymbol{K_F})}{\alpha(2\mu - \mu_1)} + \alpha\boldsymbol{E}\Phi\right\} + \\ + \boldsymbol{E}g''(\boldsymbol{K_F}).\end{aligned}$$

Substituting Expression (A.17) into the first term and simplifying yields

$$\begin{aligned}&\frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\boldsymbol{R_B}\left\{2\left(1 - \frac{1}{\alpha}\right)\frac{\boldsymbol{E}g''(\boldsymbol{K_F})}{2\mu - \mu_1} + \frac{2\boldsymbol{R_F}}{\alpha(2\mu - \mu_1)^2}\frac{\partial \mu}{\partial \alpha} + \frac{\boldsymbol{E}g''(\boldsymbol{K_F})}{\alpha(2\mu - \mu_1)} + \alpha\boldsymbol{E}\Phi\right\} + \\ &\qquad + \boldsymbol{E}g''(\boldsymbol{K_F}); \\ &\frac{\Delta^2}{\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\boldsymbol{R_B}\left\{\frac{(2\alpha - 1)\boldsymbol{E}g''(\boldsymbol{K_F})}{\alpha(2\mu - \mu_1)} + \frac{2\boldsymbol{R}_F}{\alpha(2\mu - \mu_1)^2}\frac{\partial \mu}{\partial \alpha} + \alpha\boldsymbol{E}\,\Phi\right\} + \boldsymbol{E}g''(\boldsymbol{K_F}).\end{aligned} \tag{A.19}$$

With $g'''(\boldsymbol{K_F}) > 0$ by assumption, $\alpha \geq 1$, $g''(\boldsymbol{K_F}) < 0$, $\partial\mu/\partial\alpha < 0$ and, from

Equations (A.15)-(A.16), $2\mu - \mu_1 > 0$, all terms in Expression (A.19) are negative. Hence, Condition (A.13) holds.

Part *(ii)*. Next, we prove Part *(ii)* of the Proposition. The r.h.s. of the FOC (29), given by $\boldsymbol{R_F}$ $(= g'(1 - \alpha\boldsymbol{E}))$, is increasing in α since we assume $g''(\cdot) < 0$. From the proof of Part *(i)* immediately follows that the l.h.s. of FOC (29) is decreasing in α. Then, by the Intermediate Value Theorem, there is an $\alpha \in [1, \bar{\alpha})$ that solves the FOC (29), if, at $\alpha = 1$, the l.h.s. of the FOC (29) is greater than the r.h.s. and, for $\alpha \to \bar{\alpha}$, the l.h.s. of the FOC (29) is smaller than the r.h.s. The first part of this holds true if, at $\alpha = 1$,

$$\mu_1 \boldsymbol{R_B E} + \boldsymbol{R_B E}\frac{\partial \mu}{\partial \alpha} \geq g'(1 - \boldsymbol{E})\boldsymbol{E}. \tag{A.20}$$

At $\alpha = 1$, Equation (23) yields $\hat{\boldsymbol{b}} = \Delta \boldsymbol{R_B}$. Then, Condition (A.20) holds, if

$$\begin{aligned} \mu_1 \boldsymbol{R_B E} - \boldsymbol{R_B E} R_{CB} \frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \left[\frac{\partial (p_I/p_C)}{\partial \alpha} \left(1 - \frac{1}{1}\right) + \frac{1}{1^2}\frac{p_I}{p_C} \right] &\geq g'(1 - \boldsymbol{E})\boldsymbol{E}; \\ \mu_1 \boldsymbol{R_B} - \boldsymbol{R_B} R_{CB} \frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \frac{p_I}{p_C} &\geq g'(1 - \boldsymbol{E}). \end{aligned} \tag{A.21}$$

The price ratio p_I/p_C is implicitly determined by Equation (24). For $\alpha = 1$, the solution is uniquely given by

$$\frac{p_I}{p_C} = \frac{g'(1 - \boldsymbol{E})}{\mu_1 R_{CB}}.$$

Plugging this into Condition (A.21) yields

$$\mu_1 \boldsymbol{R_B} - \boldsymbol{R_B} \frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}} \frac{g'(1 - \boldsymbol{E})}{\mu_1} \geq g'(1 - \boldsymbol{E}).$$

This holds, if

$$\mu_1^2 \boldsymbol{R_B} - \left(\mu_1 + \frac{\Delta^2 \boldsymbol{R_B}}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}}\right) g'(1 - \boldsymbol{E}) \geq 0,$$

which can be rewritten as Condition (30) in the text.

What is left to show is that the l.h.s. of the FOC (29) is smaller than the r.h.s for $\alpha \to \bar{\alpha}$. This is the case, if

$$\boldsymbol{R_B}\left[\mu - \frac{\Delta^2}{\bar{\boldsymbol{b}} - \underline{\boldsymbol{b}}} R_{CB}\left(\frac{p_I}{p_C}\frac{1}{\bar{\alpha}^2} + \left(1 - \frac{1}{\bar{\alpha}}\right)\frac{\partial(p_I/p_C)}{\partial\alpha}\right)\right] - \boldsymbol{R_F} < 0. \tag{A.22}$$

For $\alpha \to \bar{\alpha}$, the price ratio given in Equation (A.12) simplifies to

$$\frac{p_I}{p_C} = \frac{\mu_1}{2R_{CB}\frac{\Delta^2}{\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}}}\left(1 - \frac{1}{\bar{\alpha}}\right)}.$$

From Equation (A.10), the derivative of the price ratio with respect to α, "evaluated" at $\alpha \to \bar{\alpha}$, is given by

$$\begin{aligned} &\frac{\left(\frac{\mu_1(\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}})}{2\Delta^2\left(1-\frac{1}{\bar{\alpha}}\right)}\right)^2 \frac{\Delta^2}{\bar{\alpha}^2(\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}})} - \boldsymbol{E}g''(1-\bar{\alpha}\boldsymbol{E})}{R_{CB}\mu_1 - (R_{CB}\mu_1)^{(-)}} = \\ =&\frac{\frac{\mu_1^2(\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}})}{4\Delta^2\bar{\alpha}^2\left(1-\frac{1}{\bar{\alpha}}\right)^2} - \boldsymbol{E}g''(1-\bar{\alpha}\boldsymbol{E})}{0^{(+)}} = \\ =&\infty. \end{aligned}$$

It follows that the l.h.s. of Condition (A.22) goes to minus infinity, for $\alpha \to \bar{\alpha}$, and thus Condition (A.22) holds. ■

Proof of Proposition 5.[27] **Step 1.** For $\alpha \leq \alpha_m$, all banks monitor. Condition *(i)* of the proposition implies that the amount of capital provided to the BS is at inefficiently low levels for $\alpha = \alpha_m$ (i.e., $\alpha_m < \alpha_\mu$), which in turn implies that output for $\alpha < \alpha_m$ must be even lower than at $\alpha = \alpha_m$. Hence, Condition *(i)* implies that output is maximized for a value of α greater than α_m and thus for an α that implies $q = \mu < \pi$.

Step 2. We now turn to the second part of the proposition. A value of α greater than α_μ can never maximize output, since capital is already allocated efficiently and further increasing α would only decrease monitoring activity and result in an overallocation of capital to the BS. Furthermore, also a value of α equal to α_μ does not maximize output, since $\alpha^* < \alpha_\mu$ (cf. Section 5.1). Hence, if α_s and α_μ exist with $\alpha_s \geq \alpha_\mu$, output is maximized for a value of α smaller than

[27] For a more detailed version of the proof, see Gersbach and Zelzner (2022).

α_s and thus for an α that implies $q = \mu > \pi - \Delta$. The same holds true if α_μ exists but α_s does not.

Step 3. If α_μ and α_s both do not exist, or if α_s exists but α_μ does not, or if α_s and α_μ exist with $\alpha_s < \alpha_\mu$, then it is possible that output would be maximized for an $\alpha \geq \alpha_s$ and thus for an α that implies $q = \pi - \Delta$. To ensure this is not the case, we require Condition *(ii)* of the proposition. First, note that if α_μ and α_s both do not exist, output at $\alpha = \bar{\alpha}$ suddenly drops from $Y_\mu(\bar{\alpha})$ to $Y_\Delta(\bar{\alpha})$. Since α_Δ maximizes $Y_\Delta(\alpha)$, Condition *(ii)* then ensures that output stays below $Y_\mu(\alpha^*)$ for all $\alpha > \bar{\alpha}$.

Second, if α_s exists but α_μ does not, or if α_s and α_μ exist with $\alpha_s < \alpha_\mu$, then Condition *(ii)* implies $(\partial Y_\mu/\partial\alpha)|_{\alpha=\alpha_s} < 0$. We use a proof by contradiction. Assume $(\partial Y_\mu/\partial\alpha)|_{\alpha=\alpha_s} \geq 0$. Then it would be $Y_\Delta(\alpha) > Y_\mu(\alpha)$ for all $\alpha > \alpha_s$, since q is capped at $\pi - \Delta$ in $Y_\Delta(\alpha)$ but continues to decrease with α in $Y_\mu(\alpha)$. As $(\partial Y_\mu/\partial\alpha)|_{\alpha=\alpha_s} \geq 0$ would also imply $\alpha^* \geq \alpha_s$, we would obtain $Y_\Delta(\alpha^*) \geq Y_\mu(\alpha^*)$. Since α_Δ maximizes $Y_\Delta(\alpha)$, we would obtain $Y_\Delta(\alpha_\Delta) \geq Y_\mu(\alpha^*)$, which contradicts Condition *(ii)*.

From $(\partial Y_\mu/\partial\alpha)|_{\alpha=\alpha_s} < 0$ together with Condition *(i)* follows that the locally optimal leverage constraint α^* implies $\pi - \Delta < \mu < \pi$. Last, we show that Condition *(ii)* is necessary and sufficient for α^* to be the global optimum as well. As α_Δ maximizes $Y_\Delta(\alpha)$, sufficiency is obvious. To see that Condition *(ii)* is also necessary, assume it would be $Y_\Delta(\alpha_\Delta) \geq Y_\mu(\alpha^*)$. The only case in which this would not necessarily violate the existence of a globally optimal α that implies $q = \mu < \pi - \Delta$, would be the case of $\alpha_\Delta < \alpha_s$, in which $Y_\Delta(\alpha_\Delta)$ would not be "operative", since, for $\alpha < \alpha_s$, $Y_\mu(\alpha)$ applies. However, we know that $Y_\Delta(\alpha) < Y_\mu(\alpha^*)$ for all $\alpha < \alpha_s$, which, for $\alpha_\Delta < \alpha_s$, rules out $Y_\Delta(\alpha_\Delta) \geq Y_\mu(\alpha^*)$. ■

Proof of Corollary 3. Part *(i)*. Since, for $\alpha = 1$, $\hat{\boldsymbol{b}} = \Delta \boldsymbol{R_B}$, setting $\overline{\boldsymbol{b}} = \Delta \boldsymbol{R_B}$, which constitutes the first condition in the set of conditions in (32), implies $\alpha_m = 1$. Also note that, for $\alpha > 1$, it holds that $\hat{\boldsymbol{b}} < \overline{\boldsymbol{b}}\,(= \Delta \boldsymbol{R_B})$.

To see that $\alpha_s = \alpha_\Delta$, note that α_Δ is implicitly given by $(\pi - \Delta)\boldsymbol{R_B} = \boldsymbol{R_F}$.

Since $\boldsymbol{R_F} = g'(1-\alpha\boldsymbol{E})$, we can solve explicitly:

$$\alpha_\Delta = \frac{1-(g')^{-1}\left\{(\pi-\Delta)\boldsymbol{R_B}\right\}}{\boldsymbol{E}}. \tag{A.23}$$

From Condition (12), using Equation (17) and Equation (14), with $q = \pi - \Delta$, the leverage constraint beyond which all banks shirk is given by

$$\alpha_s = \frac{\boldsymbol{R_F}}{\boldsymbol{R_F} - (\pi-\Delta)\boldsymbol{R_B} + \frac{(\pi-\Delta)\underline{\boldsymbol{b}}}{\Delta}}. \tag{A.24}$$

Since by definition α_Δ implies $\boldsymbol{R_F} = (\pi-\Delta)\boldsymbol{R_B}$, α_Δ as given by Equation (A.23) falls together with α_s as given by Equation (A.24), iff

$$\frac{1-(g')^{-1}\left\{(\pi-\Delta)\boldsymbol{R_B}\right\}}{\boldsymbol{E}} = \frac{(\pi-\Delta)\boldsymbol{R_B}}{\frac{(\pi-\Delta)\underline{\boldsymbol{b}}}{\Delta}}$$
$$\underline{\boldsymbol{b}} = \frac{\Delta\boldsymbol{R_B}\boldsymbol{E}}{1-(g')^{-1}\left\{(\pi-\Delta)\boldsymbol{R_B}\right\}},$$

which constitutes the second condition in the set of conditions in (32). Furthermore, at $\alpha = \alpha_s$ it obviously holds that $\mu = \pi - \Delta$ and thus, from the definitions of α_Δ and α_μ, follows that $\alpha_s = \alpha_\Delta$ implies $\alpha_s = \alpha_\mu$.

Last, we show that the third condition in the set of conditions in (32) implies $\alpha_s \leq \bar{\alpha}$ and thus ensures the existence of α_s.[28] To see this, remember that, from Condition (A.7), $\bar{\alpha}$ is implicitly given by

$$g'(1-\bar{\alpha}\boldsymbol{E})\left(1-\frac{1}{\bar{\alpha}}\right) = \frac{\mu_1^2}{4\frac{\Delta^2}{\bar{\boldsymbol{b}}-\underline{\boldsymbol{b}}}}. \tag{A.25}$$

Furthermore, α_s is implicitly given by $\hat{\boldsymbol{b}} = \underline{\boldsymbol{b}}$, which, using Equation (23) and $p_I/p_C = \boldsymbol{R_F}/[\mu R_{CB}]$, can be written as:[29]

$$g'(1-\alpha_s\boldsymbol{E})\left(1-\frac{1}{\alpha_s}\right) = \frac{\mu}{\Delta}(\Delta\boldsymbol{R_B} - \underline{\boldsymbol{b}}).$$

As $g'(1-\alpha\boldsymbol{E})(1-1/\alpha)$ is monotonically increasing in α and as μ is monotonically

[28] In combination with the second condition, this also ensures the existence of α_μ.

[29] Note that of course $\mu = \pi - \Delta$ at $\alpha = \alpha_s$, if α_s exists. To check whether α_s exists in the first place, however, we have to take into account how μ depends on α.

decreasing in α and $\mu_1 = \pi$ for $\Delta \boldsymbol{R_B} = \overline{\boldsymbol{b}}$, it is $\alpha_s \leq \bar{\alpha}$ exactly if it holds that, at $\alpha = \bar{\alpha}$,

$$\frac{\mu}{\Delta}(\overline{\boldsymbol{b}} - \underline{\boldsymbol{b}}) \leq g'(1 - \bar{\alpha}\boldsymbol{E}) \left(1 - \frac{1}{\bar{\alpha}}\right),$$

which, with Equation (A.25), simplifies to

$$\mu \leq \frac{1}{4}\frac{\pi^2}{\Delta}. \tag{A.26}$$

From Equation (A.15) we know that $\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\Psi$ and from Condition (A.7) follows that $\Psi = 0$, at $\alpha = \bar{\alpha}$. Hence, Condition (A.26) simplifies to

$$\begin{aligned} \frac{1}{2}\pi &\leq \frac{1}{4}\frac{\pi^2}{\Delta}, \\ \Delta &\leq \frac{1}{2}\pi. \end{aligned}$$

Part *(ii)*. From Part *(i)*, $\alpha_s = \alpha_\mu$ and $\alpha_m = 1$. The former implies that increasing α beyond α^* can never maximize Y^g. This is because α^* maximizes Y^g within $\alpha_m \leq \alpha \leq \alpha_\mu$ and Y^g is certainly non-increasing for $\alpha > \alpha_\mu$, since increasing α beyond α_μ may only result in an overallocation of capital to the BS. It follows that Condition *(ii)* of Proposition 5 holds. As $\alpha_m = 1$, substituting $\overline{\boldsymbol{b}} = \Delta \boldsymbol{R_B}$ and $\mu_1 = \pi$ into Condition (30) in Proposition 4 shows that Condition (33) implies Condition *(i)* of Proposition 5. ■

References

Allen, L., Saunders, A., 1992. Bank window dressing: Theory and evidence. Journal of Banking and Finance 16, 585–623.

Benigno, P., Nisticò, S., 2022. The economics of helicopter money. URL: https://benigno.ch/wp-content/uploads/2022/03/BN20-220318.pdf.

Bloise, G., Polemarchakis, H. M., 2006. Theory and practice of monetary policy. Economic Theory 27, 1–23.

Bolton, P., Li, Y., Wang, N., Yang, J., 2020. Dynamic banking and the value of deposits. NBER Working Paper 28298.

Cavalcanti, R., Wallace, N., 1999. A model of private bank-note issue. Review of Economic Dynamics 2, 104–136.

Donaldson, J. R., Piacentino, G., Thakor, A., 2018. Warehouse banking. Journal of Financial Economics 129, 250–267.

Dubey, P., Geanakoplos, J, 1992. The value of money in a finite-horizon economy: A role for banks. In: Dasgupta, P., Gale, D., Hart, O., Maskin, E. (eds) Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn, 407–444. MIT Press, Cambridge.

Dubey, P., Geanakoplos, J, 2003a. Inside and outside fiat money, gains to trade, and IS-LM. Economic Theory 21, 347–397.

Dubey, P., Geanakoplos, J, 2003b. Monetary equilibrium with missing markets. Journal of Mathematical Economics 39, 585–618.

Dubey, P., Geanakoplos, J, 2006. Determinacy with nominal assets and outside money. Economic Theory 27, 79–106.

Faure, S. A., Gersbach, H., 2021. On the money creation approach to banking. Annals of Finance 17, 265–318.

Faure, S. A., Gersbach, H., 2022. Loanable funds versus money creation in banking: A benchmark result. Journal of Economics 135, 107–149.

Gersbach, H., 1998. Liquidity creation, efficiency, and free banking. Journal of Financial Intermediation 7, 91–118.

Gersbach, H., Rochet, J.-C., 2012. Aggregate investment externalities and macroprudential regulation. Journal of Money, Credit and Banking 44, 73–109.

Gersbach, H., Rochet, J.-C., 2017. Capital regulation and credit fluctuations. Journal of Monetary Economics 90, 113–124.

Gersbach, H., Zelzner, S., 2022. Why bank money creation? Center for Financial Studies Working Paper No. 678.

Goodhart, C., Sunirand, P., Tsomocos, D., 2006. A model to analyse financial fragility. Economic Theory 27, 107–142.

Gurley, J. G., Shaw, E. S., 1960. Money in a theory of finance. Brookings Institution, Washington, D.C.

Hahn, A., 1920. Volkswirtschaftliche Theorie des Bankkredits. J.C.B. Mohr, Tübingen.

Huber, J., Shubik, M., Sunder, S., 2014. Sufficiency of an outside bank and a default penalty to support the value of fiat money: Experimental evidence. Journal of Economic Dynamics and Control 47, 317–337.

Holmström, B., Tirole, J., 1997. Financial intermediation, loanable funds, and the real sector. Quarterly Journal of Economics 112, 663–691.

Jakab, Z., Kumhof, M., 2019. Banks are not intermediaries of loanable funds – Facts, theory and evidence. Bank of England Staff Working Paper 761.

Keynes, J. M., 1931. The consequences to the banks of the collapse of money values. In: Essays in Persuasion. Palgrave Macmillan, London.

Li, Y., Li, Y., 2021. Payment risk and bank lending. Fisher College of Business Working Paper 2021-03-017.

Macleod, H. D., 1866. The theory and practice of banking: With the elementary principles of currency, prices, credit, and exchanges, Volume 1. Longman, London.

Mcleay, M., Radia, A., Thomas, R., 2014. Money creation in the modern economy. Bank of England Quarterly Bulletin 2014Q1, 14–27.

Morris, S., Shin, H. S., 2008. Financial regulation in a system context. Brookings Papers on Economic Activity, Fall 2008, 229–274.

Parlour, C. A., Rajan, U., Walden, J., 2022. Payment system externalities. Journal of Finance 77, 1019–1053.

Piazzesi, M., Rogers, C., Schneider, M., 2022. Money and banking in a New Keynesian model. URL: `https://web.stanford.edu/~piazzesi/MBinNK.pdf`.

Schumpeter, J. A., 1954. History of Economic Analysis. Oxford University Press, New York.

Shaffer, S., Yang, S., 2010. Bank window dressing: A re-assessment and a puzzle. CAMA Working Paper 6/2010.

Shapley, L., Shubik, M., 1977. Trade using one commodity as a means of payment. Journal of Political Economy 85, 937–968.

Shubik, M., Wilson, C, 1977. A theory of money and financial institutions. Part 30 (revised). The optimal bankruptcy rule in a trading economy using fiat money. Cowles Foundation Discussion Paper 424.

Shubik, M., Tsomocos, D., 1992. A strategic market game with a mutual bank with fractional reserves and redemption in gold. Journal of Economics 55, 123–150.

Skeie, D. R., 2008. Banking with nominal deposits and inside money. Journal of Financial Intermediation 17, 562–584.

Tsomocos, D., 2003. Equilibrium analysis, banking and financial instability. Journal of Mathematical Economics 39, 619–655.

Tobin, J., 1963. Commercial banks as creators of "money". Cowles Foundation Discussion Paper 159.

Wang, T., 2019. Banks' wealth, banks' creation of money, and central banking. International Journal of Central Banking 15, 89–135.

Wang, X., 2021. Bankruptcy codes and risk sharing of currency unions. Tinbergen Institute Discussion Paper 2021-009/IV.

Wicksell, K., 1907. The influence of the rate of interest on prices. Economic Journal 17, 213–220.