Making Money: Commercial Banks, Liquidity Transformation and the Payment System*

Christine A. Parlour† Uday Rajan‡ Johan Walden§

November 16, 2017

Abstract

We consider the interaction between the dual roles of a bank as a facilitator of payments and as a lender. In our model, banks make loans by issuing digital claims to entrepreneurs, who use these claims to pay for inputs. Some digital claims are cashed in before the project is over, necessitating intermediate transfers in the interbank market. In addition, final settlement requires the lending bank to transfer reserves other banks. Each of these transfers has a cost; the endogenous interest rate in the interbank market and a settlement cost for final transfers. We consider the effects of financial innovations (i.e., FinTech) on the payment system and show that a reduction in the need for intermediate liquidity can lead to an increase in the interbank interest rate, because it also induces each bank to increase its lending. We also show that some innovation may shift investments from more productive to less productive regions and decrease productive efficiency.

*We thank Jason Donaldson, Charlie Kahn, Michael Kumhof, Michael Lee, Giorgia Piacentino, Ryan Williams, Andy Winton and seminar participants at Arizona, Boston College, Bank of Canada, the Central Bank of Ireland Microstructure Workshop, Colorado, Erasmus, FIRS, Illinois, McGill, Minnesota Corporate Finance Conference, Oxford, Tilburg, the SEC, the SED, Universidad Carlos III de Madrid, and Utah for comments.

†Haas School of Business, U.C. Berkeley, e-mail: parlour@haas.berkeley.edu.

‡Stephen M. Ross School of Business, University of Michigan, e-mail: urajan@umich.edu

§Haas School of Business, U.C. Berkeley, e-mail: walden@haas.berkeley.edu.

1 Introduction

In the US, cash is used for less than half of all consumer transactions, whereas in Norway, it is less than 5% of all transactions.[1] In both systems, the majority of consumer transactions net through the banking system, be it in the form of credit cards, debit cards or automated transfers (ACH). The promise of FinTech is to reduce the cost of making transfers between agents in the economy. But, what effect will such innovation have on the banking system? To address this question, we construct a model of a banking system in which payments are motivated by economic activity, which in turn is driven by bank lending. In our model, banks transform liquidity between deposits, reserves, loans and real output. Our particular focus is on the role of banks in the payment system, in part because this is a current focus of innovation, and in part because it is an under-investigated but pivotal function of banks.

The flows that we focus on are massive. In the United States, the Federal Funds market has had an average daily volume of U.S. $75 to $100 billion in 2017.[2] The size of the payment system, in which banks net digital claims used by their customers to make payments is even larger. This is because, in addition to funding consumption, the cash leg of all financial transactions settle through the banking system. For example, CHIPS, the largest private clearing house in the US, processed average daily volumes in 2015 of $1.436 trillion.[3] These flows are not exogenous: through their ability to make loans, and so issue demand deposits, banks have a measure of control over these magnitudes.

While off-balance sheet financing, on-balance sheet asset quality and stable funding sources have received much scrutiny, scant attention has been paid to the demand for liquidity that banks face because they provide payment and settlement services for their clients. However, ensuring a stable and efficient payment system is one of the core principles of the suggested banking supervision reforms. The Basel Committee on Banking Supervision has encouraged the adoption of new rules on liquidity risk.[4]

Understanding the role of banks in liquidity creation is somewhat complex. In the US, three currencies effectively exist in parallel and trade at par (albeit in different markets): central bank reserves, physical currency, and digital claims issued by banks. The market for central bank reserves allows banks to borrow or lend dollars, and also to effectively transfer deposits between themselves. Physical currency is typically used by households and

[1] http://www.norges-bank.no/en/Published/Publications/Norges-Bank-Papers/2017/norges-bank-papers-22017/

[2] https://apps.newyorkfed.org/markets/autorates/fed%20funds

[3] https://www.theclearinghouse.org/payments/chips/helpful-info

[4] See, for example, "Principles for Sound Liquidity Risk Management and Supervision," BIS 2008.

withdrawn from banks as part of consumer liquidity demand. Finally, digital claims that are issued by banks either as demand deposits or credit cards are used by economic agents, be they firms, entrepreneurs, or consumers, to pay for goods and services.

We construct a model of $N + 1$ strategic banks with access to investments of different quality, but with the same amount of household deposits. Each bank makes loans to a local continuum of entrepreneurs by issuing digital claims, or "fountain pen money."[5] The quantity of lending exceeds the amount of physical currency deposited at the bank. Entrepreneurs use these digital claims to purchase inputs for their technology. There are two consequences to this. First, merchants holding digital claims on a given bank may cash them in at their own local bank for cash. This interim demand for liquidity both generates a need for interbank loans and constrains the amount of digital claims banks issue compared to their cash deposits. The interest rate in the interbank loan market adjusts to clear the market. Second, as entrepreneurs pay for inputs from various merchants, banks collect digital claims issued by other banks. If, at the end of the game, one bank holds more digital claims than the other, the debtor bank transfers reserves. This is the final way in which liquidity is transformed.

In our model, the endogenous variables are the interbank rate and the amount of digital claims (inside money) issued by banks. The former is effectively the opportunity cost of real lending, while the latter corresponds to the real level of production in the economy. We take as exogenous initial household cash, merchants' demand for cash and the cost of transferring reserves through the payment system. Liquidity does not flow smoothly in our model because of three basic frictions. First, banks face a deadweight cost to being a net payer in the payment system. This cost comprises both the collateral cost of using the current systems and the internal opportunity cost of liquidity management. The importance of this friction is evinced by the Basel Committee's focus on liquidity management and the implementation of regulations that require a bank to hold liquid assets in proportion to its projected net outflows.[6] The cost associated with being a net payer reduces a bank's willingness to issue digital claims. The second friction is a merchant's intermediate liquidity need, which requires funds to be transferred in the interbank market in anticipation of this demand. The equilibrium interest rate in the interbank market acts as an endogenous friction in the sense that the most productive banks have to share some of the surplus from investment, which reduces their willingness to invest. The third friction is a strategic one: banks with similar productivity want to mimic what other banks are doing so that they do not have to

[5] This (somewhat dismissive) phrase is due to Tobin (1963).

[6] See, for example, the discussion in "Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools", BIS, January 2013.

incur the cost of transferring payments.

We characterize two classes of equilibria that differ in the amount of digital claims (inside money) that banks issue. Symmetric equilibria occur if local conditions between banks are sufficiently similar, whereas asymmetric ones occur when local conditions are sufficiently heterogeneous. In economies in which the productivity of banks is sufficiently similar, a coordination friction manifests itself. Banks know that they will face a cost to being a net payer in the payment system and so rationally issue only as many claims as the average bank in the economy. Because of the coordination friction, the interbank rate can only be pinned down within some range. By contrast, if the economy features sufficiently heterogeneous banks, the interbank rate is uniquely determined. In such economies, more productive banks are net borrowers in the interbank market and net payers through the payment system. The banking system is therefore a device for transferring resources from less productive to more productive banks.

We next turn to the effect of financial innovation (i.e., FinTech). Our model is a partial equilibrium one, and so we do not focus on welfare. Rather, because bank lending affects real output, we focus on productive efficiency. Innovations that result in increases in real output are deemed more efficient. We focus on two kinds of FinTech innovations: Those that affect consumers, and those that affect back-office banking operations.

Suppose we move to a world in which new methods are developed for consumer-to-consumer payments. This will have two effects: The demand for cash decreases, and the degree of integration across banks increases. There are two parameters in our model that correspond to such changes. We establish that a decreased consumer demand for cash will lead to banks issuing more digital claims (so an increase in the money multiplier) and can lead to an increase or a decrease in the interbank rate. However, a decrease in demand for cash will unambiguously increase the productive efficiency of the system. By contrast, if the economy become more integrated, there will be a weak increase in the interbank rate. This adversely affects more productive banks, and investments shift from more productive to less productive regions. Productive efficiency decreases. The equilibrium effect of innovation in consumer payment methods on the banking system is therefore nuanced.

The other source of FinTech innovation that has received attention is clearing and settlement. Specifically, the move to a distributed ledger technology (such as the innovations discussed at SWIFT or the Faster Payments initiative) will reduce the need for collateral and therefore the overall cost of transferring money. We show that efficiency increases in the interbank payment system will, somewhat counter-intuitively, lead to increases in the interbank rate. The cost of transferring funds currently falls on the most productive banks.

If this cost falls, these banks issue a greater quantity of digital claims, and therefore increase their investment. A side effect of this will be an increase in the interbank rate.

Modeling details aside, our analysis has three important elements. First, banks facilitate payments which generate large, predictable interbank flows. Second, payment flows are tied to the level of demand deposits held by banks. Some demand deposits originate from households; however, banks also create demand deposits when they issue loans. Third, banks are strategic and take these flows into account. Thus, our work is related to the literature that micro-founds the actions of banks and the macro literature that examines the conduct of monetary policy.

Our paper is most closely related to those on liquidity management in the banking system. A very early paper in this vein is Edgeworth (1888). More recently, Bianchi and Bigio (2014) present a calibrated general equilibrium model in which banks receive deposits, make loans and settle reserves in the interbank market. They highlight the importance of credit demand shocks. By contrast, in our partial industry analysis, we focus on the effect of strategic credit supply and emphasize the importance of the payment system in generating interbank flows.

Bank liquidity is important in part because it can affect bank solvency. This feature is considered in Freixas and Parigi (1998) and Freixas, Parigi, and Rochet (2000), both papers that consider aspects of contagion and risk spread through the payment system. Plantin (2015) presents a model in which bank insolvency is socially costly because agents use bank claims to settle transactions. Our model features no risk of bank solvency; instead, we focus on non-bankruptcy-related frictions that emerge from the payment system.

Some researchers have started from the premise that agents value liquid assets, and have considered banks' (and others') incentives to produce such assets (see, for example, Gorton and Pennacchi, 1990). Gu, et al. (2013) micro-found the existence of banks as both payment facilitators and deposit takers. They emphasize the importance of commitment and show that allocations are inferior without banks.

Our analysis differs from the seminal work of Diamond and Dybvig (1986) in two fundamental ways. First, as we are interested in frictions other than bank insolvency, our banks face no project risk. Second, in the Diamond and Dybvig world, banks are simply a conduit to channel savings from households to the real economy. We allow banks to issue fountain pen money, which implies that banks must actively choose the quantity of loans they make.

The microfoundations of payment systems and, in particular, the costs and benefits of gross versus net settlement are examined by Kahn and Roberds (2009b), Kahn, McAndrews, and Roberds (2003), Kahn and Roberds (1998) and Bech and Garratt (2003). Some of this work considers the strategic interaction between banks as they process payments. Our view is somewhat longer as we consider the horizon over which banks also issue demand deposits.

After the financial crisis, there has been renewed interest in the interbank market in which banks borrow and lend reserves. Ashcraft, McAndrews and Skeie (2009) present a detailed empirical analysis of the interbank market, and present evidence that large banks typically borrow from small banks and do so as a result of liquidity shocks that arise because of large value transfers. Recent theoretical work including Acharya, Gromb and Yorulmazer (2012) considers various aspect of the interbank market including market power on the side of banks who do not face liquidity shocks. We differ from this literature in that, in order to focus on how the source of the liquidity shocks affects bank behavior.

The history of early banking focuses on intermediaries' role in transferring value and providing a payment system. See, for example, the descriptions in Kohn (1999) and Norman, Shaw and Speight (2011). Both papers mention the extensive Medieval fair at Champagne in which trading was divided into the early sale of cloth and the later sale of spices. Merchants from Flanders sold cloth which was purchased by the Italians, whereas the spice importers were the Italians who sold to the merchants from Flanders. Payments were effectuated by transfers of credits through banks, and were not necessarily backed by cash coins. In effect, the banks facilitated a complicated barter arrangement between differing pairs of traders. This arrangement mirrors a net settlement system. Kahn and Roberds (2009a) provide an introduction to the economics of payment and settlement systems in the modern economy.

As our banks operate across two markets (the local and foreign market), the structure of the economy is similar to a two-sided market as surveyed in Rochet and Tirole (2006). Competition for deposits and the non-existence of two sided Bertrand equilibria are presented in Yanelle (1997), which justifies our assumption that banks have market power. Donaldson, Piacentino and Thakor (2015) consider the features of banks that make them the optimal entity to issue chits; i.e., digital claims.

2 Model

An economy comprises $N+1$ banks, each of which is indexed by i. Each bank i may be thought to represent a local banking system. Banking systems are distinct from each other as each has its own segmented markets. This could arise from either geographic differentiation or product differentiation along some unspecified dimension. We often refer to banking system i as zone i. Each bank i accepts deposits, makes loans to entrepreneurs, and facilitates payments. Associated with each bank are a continuum of identical entrepreneurs and a continuum of identical merchants, so each bank may be thought of as a banking system. We frequently refer to merchant i or entrepreneur i; these are the typical entrepreneur or merchant that banks with bank i. The size of each of these continua is normalized to 1.

Two media of exchange circulate between agents and banks: (i) central bank money, that can take the form of fiat money ("cash") or reserves that we denote by z, and (ii) digital claims issued by commercial banks when they make loans, denoted by d (these are issued to entrepreneurs as demand deposits). Although not formally included in the model, there is implicitly a central bank which issues fiat money and holds commercial bank reserves, and which can convert reserves into cash.

Following practice in the U.S., in our model cash, digital claims, and central bank reserves are all denominated in the same units (dollars) and are exchangeable at par from one form to another. As will be clear shortly, there is no danger of bank insolvency, so no need to discount digital claims. Therefore, each unit of d and z is price at \$1. There are also real goods or inputs (k) that are converted by the entrepreneur's technology into real output. In our base model, the price of one unit of real inputs in each zone is normalized to \$1. The implicit assumption here is that the supply curve for these inputs is flat at the price of \$1.

Liquidity transformation in this model takes place over four dates, $t = 0, 1, 2, 3$. Final payoffs are realized at date 4. The timeline is presented in Figure 1.

At time $t = 0$, local households deposit cash in their bank. Because we are interested in the effect of productivity differences across banks on liquidity creation and investment, we assume that each bank receives the same level of deposits, D. Each bank i then makes loans in the form of take-it-or leave it offers to its local entrepreneurs. It does so by creating demand deposits d_i. Typically, the loans extended, or "inside money," exceed the cash deposits — as discussed later, entrepreneurs' projects can generate a surplus, creating the resources necessary to repay the higher loan amounts.

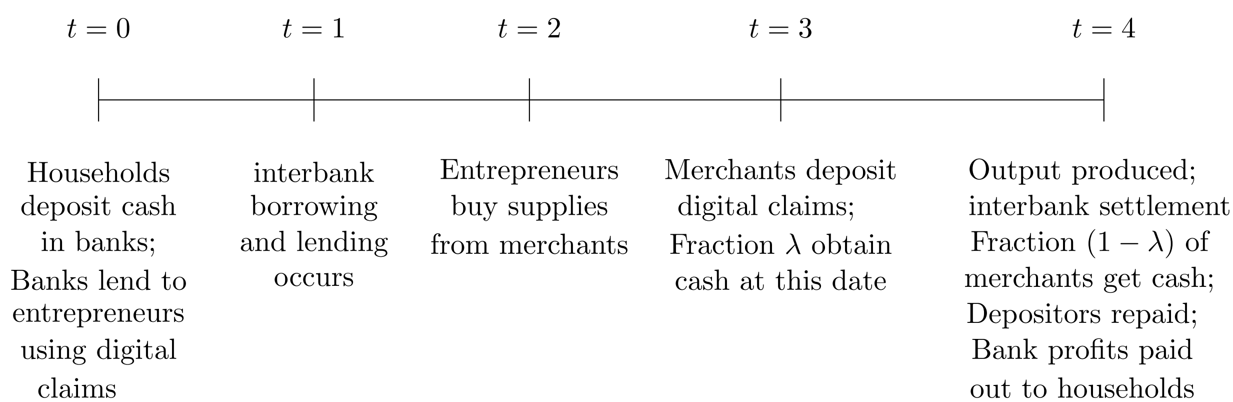

Figure 1: **Sequence of events.**

At $t = 1$, the interbank market is active. This market is akin to the federal funds market, and allows banks to trade reserves. Reserves can be exchanged between banks and also transformed by them into cash and paid out. This is the market for interim liquidity and we denote the interest rate (i.e., the price of borrowing and lending reserves) at which the market clears as r_b. Banks trade reserves in anticipation of their liquidity needs at date 3.

At $t = 2$, entrepreneurs use digital claims to purchase their inputs from merchants. Each entrepreneur in zone i has access to a production technology. With k units of real inputs, the technology in zone i produces $\mu_i f(k)$ of the consumption good. Here, $f(k)$ is a strictly increasing and strictly concave production function common to all zones, and μ_i is a zone-specific productivity parameter. Without loss of generality, we assume that productivities are ordered, so that $0 < \underline{\mu} = \mu_1 \leq \mu_2 \cdots \leq \mu_{N+1}$. That is, the entrepreneurs associated with bank 1 have the lowest productivity in the economy, denoted by $\underline{\mu}$. We also assume that the difference in productivity between any two banking zones is constant, at $\mu_{i+1} - \mu_i = \delta \geq 0$. This assumption will be useful to characterize equilibria.

Let d_i denote the size of digital claims available to entrepreneur i, and k_i the quantity of inputs she purchases from merchants. Merchants are perfectly competitive, and the input price is normalized to 1, so that $k_i = d_i$. Following Freixas and Parigi (1998) and Freixas, Parigi and Rochet (2000), we introduce the payment system by assuming that entrepreneurs are randomly assigned to buy inputs in different locations. Specifically, entrepreneur i needs $(1-\alpha)k_i$ inputs locally and αk_i inputs from other zones, where $\alpha \in [0, 1]$. For tractability, we assume that the demand for foreign inputs is equally divided among zones, so the entrepreneur purchases $\frac{\alpha k_i}{N}$ from each foreign zone. As α increases, the level of integration across zones increases. Autarky is represented by $\alpha = 0$, and when $\alpha = \frac{1}{N+1}$, the entrepreneur's input needs are evenly divided across the entire economy.

At $t = 3$, merchants deposit the various digital claims they have received (from entrepreneurs in different zones, and hence drawn on different banks) in their own bank. Collectively, each entrepreneur's purchase of inputs generates liquidity demand from the payment system at each merchant's bank. Specifically, a proportion $\lambda \in [0, 1]$ of merchants are impatient, and have a need for consumption before the output is realized. These merchants arrive at their bank at date 3 and immediately withdraw their digital claims. To consumer, they need to be provided with central bank money — either cash or a claim on the central bank. Each bank i therefore needs to either hold cash reserves (obtained at date 0 from households) or must borrow reserves from other banks (at date 1) to meet this liquidity demand. A proportion $1 - \lambda$ of merchants are patient, and do not consume at the intermediate date. They hold their digital claims with their own bank, and convert them into cash only after the output is realized.

At $t = 4$, production is realized, final interbank settlements are made, and final remittances are made to merchants and households. When the output is produced, entrepreneurs deposit it back into their own bank, and the bank makes all payments. If $\alpha > 0$, then interbank settlement may be required. In this case, bank i has issued some digital claims that have been turned in by merchants in zone j to bank j. These claims are still liabilities of bank i, so bank i must transfer reserves to bank j to fulfill these claims. We assume that only net (rather than gross) amounts are settled between banks. That is, in this market for ex post liquidity, banks that are net debtors transfer reserves to banks that are net creditors in the payment system.

There is a cost of transferring funds between banks, which we denote by τ, for final settlement at date 4. This transfer cost plays an important role in our analysis, and represents three underlying frictions. First, under the Basel III Accords, the liquidity coverage ratio is defined based on a bank's future net cash outflows.[7] To the extent a bank has to hold liquid assets in anticipation of future cash outflows, it cannot lend these funds out for long-term projects, giving rise to an opportunity cost. Second, many payment systems require banks to pre-deposit collateral with the system before they can participate. Further, the amount of collateral a bank needs is in proportion to its net outflows. For example, in the US, the CHIPS system processes \$1.5 trillion in cross-border and domestic payments daily, and requires the participant banks to post collateral based on their historical transfer volume before they can participate in the system.[8] Collateral used to participate in the payment system cannot be used for other purposes (such as raising funds in the repo market), and so a second opportunity cost arises. Third, most payment systems, including FedWire and CHIPS, have explicit (albeit small) fees that must be paid for using the system.

Finally, at date 4, remittances are made to merchants and households. A proportion $1 - \lambda$ of merchants convert their digital claims into cash at this date. The payment to households has two components. First, deposits are returned; we normalize the interest rate on deposits to zero. Second, households own the bank, and thus receive all profits generated by the bank.

We impose two parameter restrictions. The first is a restriction on the marginal productivity of the least productive bank. The condition ensures that the bank would still invest all

[7]Specifically, the denominator is the total net cash outflows of the bank over the next 30 days. See "Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools," available at http://www.bis.org/publ/bcbs238.pdf .

[8]See `https://www.theclearinghouse.org/payments/chips/helpful-info` and `https://www.newyorkfed.org/aboutthefed/fedpoint/fed36.html` for more information.

its resources even under autarky. The second restriction ensures that the cost of interbank transfers through the payment system, τ, is not prohibitively large.

Assumption 1 *The economy satisfies:*

(i) High minimum productivity: $\underline{\mu} f'(\frac{D}{\lambda}) - 1 > 0$

(ii) Small transaction costs: $\tau < \frac{1}{N\alpha}$.

2.1 Liquidity Flows in the Economy

The liquidity flows for one zone are presented in Figure 2. The flows are somewhat involved. There are a few different financial claims in this market. In the figure we distinguish between physical cash and central bank reserves; in the model, we treat these claims as equivalent. Claims issued by bank i are referred to as digital claims. Eventually, all claims are claims on the consumption good. Entrepreneurs pay merchants with digital claims, and receive real inputs. Merchants deposit claims in banks, and receive cash to purchase the consumption good. In the market for interbank liquidity (the "interbank market"), banks trade reserves at date 1. At date 4, all settlement payments on interbank loans and digital claims are made.

Consider the relationship between digital claims issued by a bank and its demand for ex ante liquidity. Suppose that banks in the economy have issued digital claims $d_1, \ldots d_{N+1}$, which entrepreneurs have used to buy inputs. Then, the digital claims received by merchant i in exchange for his real inputs are:

$$S_i = (1-\alpha)d_i + \frac{\alpha}{N}\sum_{j\neq i} d_j. \tag{1}$$

The first term comes from entrepreneurs in zone i who have purchased inputs locally, and the second term reflects the supplies sold by merchant i to entrepreneurs in other zones $j \neq i$.

Merchant i deposits these claims in his local bank at date 3. With probability λ, he is hit by a liquidity shock and withdraws cash at this date. Thus, to remain solvent, the bank needs to have enough cash on hand to satisfy this interim demand, which amounts to λS_i. As S_i depends in part on claims issued by other banks, the interim demand λS_i may be either more or less than the cash deposits D that bank i holds. Define $z_i = \lambda S_i - D$. If $z_i > 0$, the bank borrows current cash in the interbank market. Conversely if $z_i < 0$, the surplus can be lent out in the interbank market. Interbank loans are settled at date 4, leading to a cash flow for bank i of $-(1+r_b)z_i$ at this date. We assume transfers across banks related to interbank loans are costless both at date 1 and at date 4, as they correspond to book-keeping adjustments.

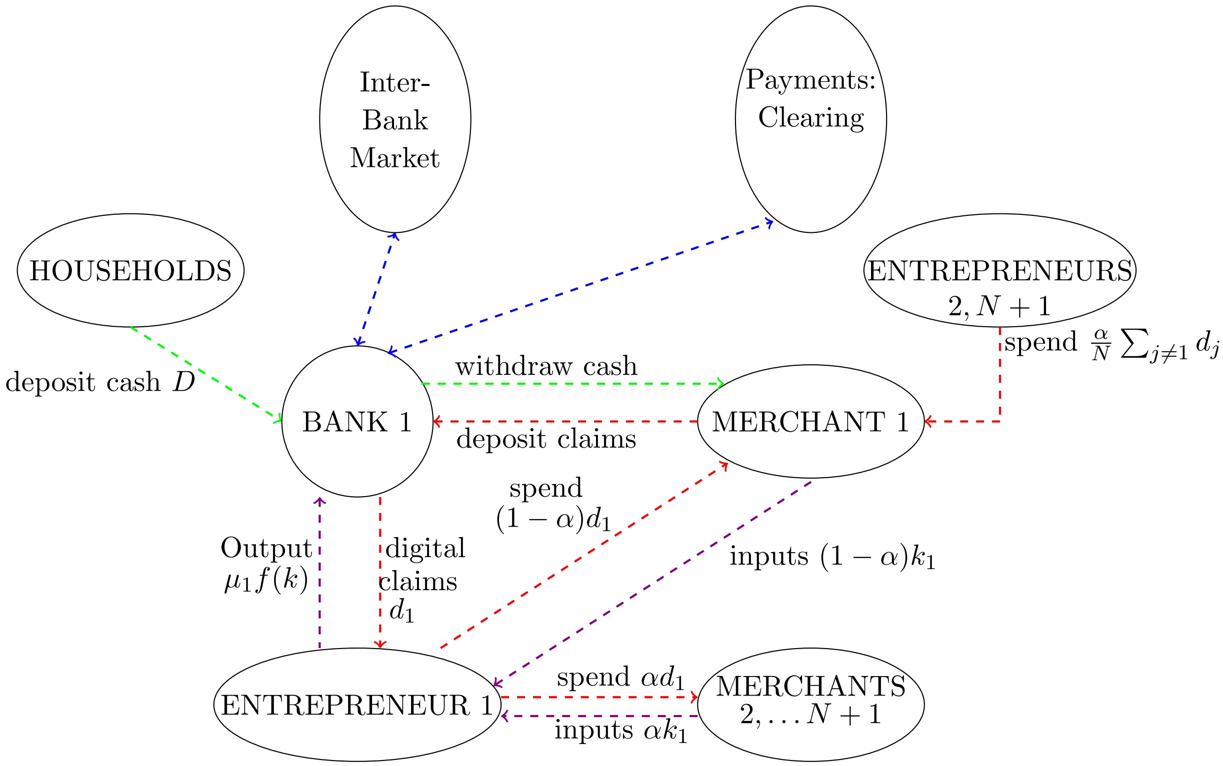

Figure 2: **Liquidity transformation in the economy.** The figure shows the flows of deposits, cash reserves, and digital claims for Zone 1. Cash flows are in green, digital claims are in red, real goods are in violet, and reserves are in blue.

At the settlement stage at date 4, ex post liquidity is transferred across banks. At this stage, the banks net out the digital claims that are circulating in the economy. The total number of claims issued by bank i that are held by other banks are αd_i, while bank i holds $\frac{\alpha}{N}\sum_{j\neq i} d_j$ from other banks. Hence, the net payment outflow amount for bank i at date 4 as a result of this effect is

$$\alpha d_i - \frac{\alpha}{N}\sum_{j\neq i} d_j. \tag{2}$$

This amount too may be positive or negative, and represents the amount bank i has to transfer to other banks as part of its role in the payment system. If the amount is positive, bank i has to pay an additional cost of τ per unit to effect the transfer. Notice that the transfer fee is asymmetric — a sending bank pays it, but a receiving bank does not earn it. This is consistent both with τ representing an opportunity cost for the paying bank and with the fact that typically, fees and costs associated with payment systems are paid to third parties.

2.2 Bank Objective Function

Each bank seeks to maximize its profit, by choosing how many digital claims to issue. Profit is denominated in units of the consumption good at date 4, and has three components. First, the bank obtains the surplus from production, and entrepreneurs are held down to their reservation utility, which is normalized to zero.[9] Suppose the bank issues d_i digital claims. As claims are converted at a price of 1 into real inputs, the quantity of inputs purchased, k_i, equals d_i. Therefore, the expected surplus from production is $\mu_i f(k_i) - k_i = \mu_i f(d_i) - d_i$. Note that of the d_i claims eventually held by merchants, a proportion λ are paid at date 3, and the remainder are paid at date 4. There is no discounting, so the cumulative value of these claims at date 4 remains d_i.

The second component of the profit function is due to flows in the interbank lending market at date 1. Bank i has a net inflow of $z_i = \lambda\left((1-\alpha)d_i + \frac{\alpha}{N}\sum_{j\neq i} d_j\right) - D$ at this date. At date 4, it must transfer to its lenders $z_i(1+r_b)$. If $z_i < 0$, both flows are in the opposite direction. Again using the fact that there is no discounting, the net value of these flows at date 4 is $r_b z_i$; that is, the interest cost or gain from borrowing or lending in the interbank market.

Finally, at date 4, bank i makes or receives a net transfer of funds to other banks for final settlement of its own digital claims. As shown in equation (2) above, the size of this transfer is $\alpha d_i - \frac{\alpha}{N}\sum_{j\neq i} d_j$. If this quantity is positive, the bank incurs a cost τ per unit. Therefore, the cost of this transfer is $\tau\left(\alpha d_i - \frac{\alpha}{N}\sum_{j\neq i} d_j\right)^+$, where for any variable x, $x^+ = \max\{x, 0\}$.

Putting all this together, bank i's payoff function is

$$\pi_i = \mu_i f(d_i) - d_i - r_b z_i - \tau\alpha\Big(d_i - \frac{1}{N}\sum_{j\neq i} d_j\Big)^+ \qquad (3)$$

The bank faces an interim solvency constraint at date 3 — it must have enough cash on hand to meet merchant needs at that date. The demand for liquidity at bank i amounts to $\lambda\left(\alpha d_i + \frac{1-\alpha}{N}\sum_{j\neq i} d_j\right)$. The supply of liquidity can come from two sources: deposits, D, and interbank borrowing or lending, z_i. Note that we assume that interbank loans are immediately available as cash to the borrowing bank; that is, we treat central bank reserves, which are claims on the central bank, as equivalent to physical cash. The interim liquidity

[9] There is no competition between banks for entrepreneurs, merchants or depositors. We implicitly appeal to unmodeled taste differences to generate such segmentation.

constraint at date 3 is therefore

$$z_i + D \;\; \geq \;\; \lambda\Big((1-\alpha)d_i + \frac{\alpha}{N}\sum_{j\neq i} d_j\Big). \tag{4}$$

It is easy to see that the interim liquidity constraint (4) is binding. Suppose that it is slack, so that the bank is holding more cash than it needs to satisfy merchants' demand. In this case, if $r_b > 0$, a bank could simply lend out its extra reserves in the interbank market and increase profits. Thus, constraint (4) must hold as an equality. That is,

$$z_i \;\; = \;\; \lambda\Big((1-\alpha)d_i + \frac{\alpha}{N}\sum_{j\neq i} d_j\Big) - D. \tag{5}$$

Substituting this into the bank's objective function, (3), we have

$$\pi_i \;\; = \;\; \Big(\mu_i f(d_i) - d_i\Big) - \tau\,\alpha\left(d_i - \frac{1}{N}\sum_{j\neq i} d_j\right)^{+} - r_b\,\lambda\left((1-\alpha)d_i + \frac{\alpha}{N}\sum_{j\neq i} d_j\right) + r_b D. \tag{6}$$

The interbank rate r_b as given, it is an important equilibrium quantity, and is determined by market-clearing in the market for reserves at date 1; i.e., by the requirement that $\sum_{i=1}^{N+1} z_i = 0$. Summing both sides of equation (5) over the number of banks in the economy, this condition may equivalently be expressed as

$$\underbrace{\lambda\sum_{i=1}^{N+1} d_i}_{\text{Demand for date 3 liquidity}} \qquad = \qquad \underbrace{(N+1)D}_{\text{Supply of liquidity}} \quad . \tag{7}$$

As we will show below, a bank's optimal quantity of digital claims, d_i, depends on r_b. Further, because of the interbank market, while the volume of household deposits does not directly determine the volume of loans that any single bank makes, in the aggregate the quantity of digital claims issued does depend on the amount of deposits (or cash) available in the economy.

3 Equilibrium

We characterize a Nash equilibrium in the banks' lending game, with each bank taking the interest rate in the interbank market, r_b, as given. Banks have complete market power in their own zone (given that there is a continuum of entrepreneurs and no bank lending across zones), and thus are strategic in the issuance of digital claims. Our implicit assumption is that even the most productive bank is sufficiently small in the market for reserves to take the interest rate as given.

The interest rate r_b, in turn, clears the interbank market at date 1.

Definition 1 *An equilibrium in the model consists of a vector of digital claims* $(d_1^*, \cdots, d_{N+1}^*)$ *and an interest rate in the interbank market,* r_b^**, such that:*

(i) For each bank i*,* d_i^* *maximizes its payoff* π_i*, given* r_b^* *and the digital claims issued by all other banks,* $(d_1^*, \cdots, d_{i-1}^*, d_{i+1}^*, \cdots, d_{N+1}^*)$*.*

(ii) The interbank loan market clears; that is, $\sum_{i=1}^{N+1} d_i^* = \frac{(N+1)D}{\lambda}$*.*

3.1 Bank i's Best Response

At date 0, each bank i chooses d_i, the number of digital claims it issues, taking as given the number of claims issued by every other bank and the interest rate in the interbank market at date 1. As entrepreneurs are all identical, each entrepreneur in zone i receives the same number of claims. Recall that there is a mass 1 of entrepreneurs; we can therefore refer to d_i as the number of claims received by a single entrepreneur in zone i.

Recall that the inputs purchased by entrepreneur i equal d_i, because the input price is normalized to 1. Therefore, we can equivalently think of d_i as the size of the real investment made by entrepreneurs in zone i. For clarity, we refer to k_i as the investment and d_i as the number of digital claims, with $k_i = d_i$. It is important to keep in mind, therefore, that the quantity of digital claims corresponds with the amount of real activity in the economy.

Define $h(x) = f'^{-1}(x)$. That is, $h(\cdot)$ recovers the real input level that generates a particular level of marginal product. We note that the concavity of $f(\cdot)$ implies that $h(x)$ is decreasing in x—higher marginal products are generated by lower input levels.

Two thresholds are useful in exhibiting the best response function of bank i. Define

$$\hat{d}_i = h\left(\frac{1 + r_b\lambda(1-\alpha)}{\mu_i}\right) \tag{8}$$

$$d_i^\tau = h\left(\frac{1 + r_b\lambda(1-\alpha) + \tau\alpha}{\mu_i}\right). \tag{9}$$

That is, $\hat{d}_i$ denotes the investment or input level at which the marginal product in zone i is equal to $\frac{1+r_b\lambda(1-\alpha)}{\mu_i}$, and d_i^τ the input level at which the marginal product in zone i is equal to $\frac{1+r_b\lambda(1-\alpha)+\tau\alpha}{\mu_i}$. As $h(\cdot)$ is decreasing, it is immediate that $\tau > 0$ implies that $\hat{d}_i < d_i^\tau$. Note that both $\hat{d}_i$ and d_i^τ are functions of r_b; where convenient, we refer to $\hat{d}_i(r_b)$ and $d_i^\tau(r_b)$.

These thresholds are illustrated in Figure 3 below. Fix the claims issued by banks $j \neq i$, and the interest rate in the interbank market at date 1, r_b. If bank i issues an extra digital claim, it faces two liquidity-related effects. First, a proportion $\lambda(1-\alpha)$ of these claims will be cashed in at bank i by merchants i at date 3. This incurs an additional borrowing cost

$r_b\lambda(1-\alpha)$, which shows up in the numerator of both $\hat{d}_i$ and d_i^τ. Second, if bank i has issued more claims than the average across all other banks, it has to transfer additional resources to other banks at date 4, to pay off all digital claims of bank i now held by the other banks. If bank i issues one extra digital claim, the cost of this additional transfer at date 4 is $\tau\alpha$, the last term in the numerator of d_i^τ. To overcome these extra liquidity costs, in order to issue another digital claim, it must be that the productivity of an entrepreneur's project $(\mu_i f'(d_i))$ is higher than the marginal cost of the input (fixed at 1) and the transfer cost.

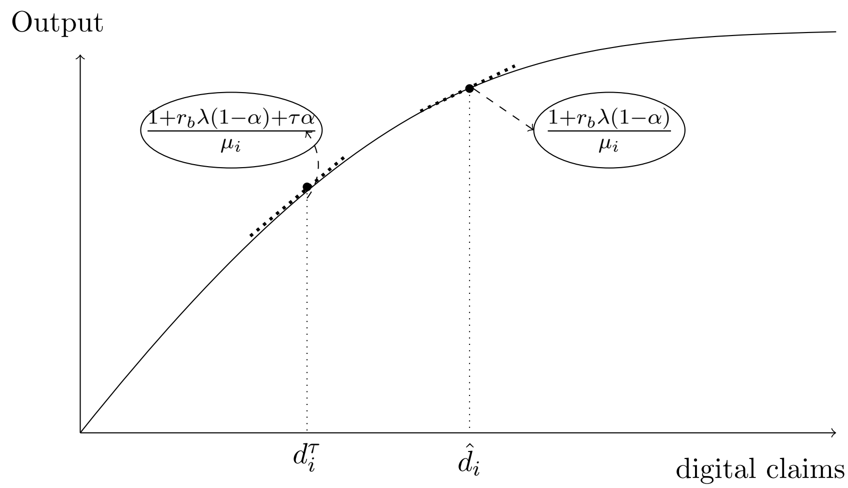

Figure 3: **Critical output thresholds when** $\tau > 0$. At d_i^τ, the marginal product includes the transfer cost, τ. This corresponds to a lower level of claims and inputs than at $\hat{d}_i$.

Bank i's best response depends in part on the amount it expects to transfer to other banks at date 4. As all zones are symmetric in terms of the proportion of foreign inputs sought by their entrepreneurs (i.e., α is common across all zones), the best response depends only on the average number of digital claims issued by other banks, $\frac{1}{N}\sum_{j\neq i} d_j$. The positive transfer cost τ creates a coordination friction—there is a range for the average claims issued by other banks such that bank i wishes to exactly match this average. If it matches the average, it neither makes nor receives any transfers from other banks, either in the interbank market at date 1, or in the final settlement at date 4. If, instead, it choses a higher level of digital claims d_i, it is subject to additional transfer costs at date 4. Conversely, at a lower level of digital claims, the marginal profit is still increasing, because a small increase in digital claims imposes no additional transfer cost at date 4. This feature arises because of the asymmetry in the transfer cost τ: A debtor bank incurs this cost, which acts as a deadweight friction, rather than being paid to a creditor bank.

Lemma 1 *Suppose banks $j \neq i$ issue aggregate digital claims $\sum_{j\neq i} d_j$. Then, the best response of bank i is*

$$d_i^* = \begin{cases} d_i^\tau & \text{if } d_i^\tau \geq \frac{1}{N}\sum_{j\neq i} d_j \\ \frac{1}{N}\sum_{j\neq i} d_j & \text{if } \frac{1}{N}\sum_{j\neq i} d_j \in (d_i^\tau, \hat{d}_i) \\ \hat{d}_i & \text{if } \hat{d}_i < \frac{1}{N}\sum_{j\neq i} d_j \end{cases} \tag{10}$$

We begin by observing that, in any equilibrium, more productive banks issue weakly more claims. It is important to keep in mind throughout that the input level in zone i, k_i, equals the number of digital claims issued. Therefore, the number of digital claims corresponds directly to the level of real activity in the economy.

Lemma 2 *In any equilibrium, $d_1^* \leq d_2^* \leq d_{N+1}^*$. That is, the number of digital claims issued is weakly greater in more productive zones.*

We consider two kinds of equilibria. First, we look at symmetric equilibria, in which all banks choose the same level of digital claims. These equilibria occur if zones are sufficiently similar in their productivity. In such equilibria, all local deposits are used to fund local projects, and there is a limited role for banks to re-allocate funds through the interbank market. Next, we consider asymmetric equilibria, which occur when the productivities across zones are sufficiently different. In these equilibria, more productive zones issue a strictly greater number of digital claims than less productive zones. As each zone starts out with the same level of deposits, the interbank market must re-allocate funds across zones.

Our framework also admits hybrid equilibria in which a cluster of low-productivity banks issue the same number of claims as each other, with more productive banks separating out and issuing more claims. The properties of such equilibria are directly discernible from the two extreme cases that we consider, so going forward we do not discuss these equilibria.

First consider symmetric equilibria. In this case, there are no net flows in the interbank market or the net payment settlement system. Such equilibria occur if all zones have similar productivities. In this world, the interbank rate is still defined, and adjusts so that banks do not want to borrow or lend. Thus, even though there is no volume of trade in the interbank market, the market price is still defined by market clearing.

Define

$$\hat{\delta} = \frac{\tau\alpha}{Nf'(\frac{D}{\lambda})}. \tag{11}$$

If $\delta \leq \hat{\delta}$, productivity is approximately homogeneous across zones. In these sufficiently homogeneous economies, in equilibrium each bank lends the same amount as it would under autarky. That is, $d^* = \frac{D}{\lambda}$. Surprisingly, however, the coordination friction implies that the price that clears the interbank market is not necessarily unique, but will lie in a range.

Proposition 1 *Suppose $\delta \leq \hat{\delta}$, so that the productivity difference across zones is sufficiently small. Then, for each $r \in \left[\frac{(\underline{\mu}+\delta N)f'(\frac{D}{\lambda})-(1+\tau\alpha)}{\lambda(1-\alpha)}, \frac{\underline{\mu}f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)}\right]$, there is an equilibrium in which the interest rate in the interbank market is given by $r_b^* = r$. Further, in each such equilibrium:*

(i) Each bank issues the same quantity of digital claims, with $d_1^* = d_2^* = \cdots = d_{N+1}^* = \frac{D}{\lambda}$.

(ii) At date 4, net payment flow between any two banks is zero.

Even though no banks have a net debt through the payment system, the existence of an ex post settlement cost, τ, generates the interval of prices. If $\tau = 0$, then the bank's problem has a unique solution, or $d^\tau = \hat{d}$. In this case, there would be a unique interbank market clearing price of $\frac{\underline{\mu}f'(\frac{D}{\lambda})-1)}{\lambda(1-\alpha)}$. The coordination friction that obtains for $\tau > 0$ translates into a range of prices when the interbank market is open. The indeterminacy in the market clearing price could manifest itself as either volatility or stickiness in the interbank market. In particular, observed changes in the interbank rate do not necessarily correspond to changes in lending activity or inside money creation.

It is important to recognize that the autarkic production outcome is not efficient, because projects in different zones have different productivities (i.e., μ parameters). Intuitively, aggregate expected output is highest if the expected marginal product of capital is equalized across zones. As we show below in the heterogeneous productivity case, in general two frictions prevent achieving the efficient outcome: the cost of making settlement payments at date 4, τ, and the cost of borrowing in the interbank market at date 1, r_b. The same two frictions operate with sufficiently homogeneous productivity as well.

Observe that, if τ and r_b were both set to zero, in a Nash equilibrium in lending we would have $\mu_i f'(d_i^*) = \mu_j f'(d_j^*)$ for every pair of banks i and j. In other words, the marginal productivity of investment would be equalized across zones. With both homogeneous and heterogeneous productivity across zones, actual investment departs from this frictionless benchmark.

Of course, one of the important functions of capital markets is to reallocate funds from less productive to more productive areas. We now turn our attention to an economy in which different banks serve entrepreneurs with projects of sufficiently heterogeneous quality. In other words, consider the case in which δ is sufficiently high. As shown in Proposition 1, when δ is low, a coordination friction exists across banks. To ensure that no two banks i and $i+1$ have the same investment level, we require δ to exceed a minimum threshold. Define

$$\underline{\delta} = \frac{\underline{\mu}\tau\alpha}{\underline{\mu}f'(\frac{D}{\lambda}) - (N-1)\tau\alpha}. \tag{12}$$

We show that when the productivity difference across consecutive zones, δ, exceeds this threshold, investment in more productive zones exceeds that in less productive zones.

Lemma 3 *Suppose that $\delta > \underline{\delta}$. Then, in any equilibrium,*

(i) $r_b^* > \frac{\mu f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)}$.

(ii) $d_1^* < d_2^* < \cdots < d_{N+1}^*$.

Further, when productivities are sufficiently heterogeneous, there is a unique equilibrium. While the aggregate investment in the economy remains $\frac{(N+1)D}{\lambda}$, the distribution of investment across banks varies. Therefore, the total output in this equilibrium is greater than under autarky.

Proposition 2 *Suppose that $\delta > \underline{\delta}$, so that the productivity difference across zones is sufficiently large. Then, in equilibrium, there is a unique market-clearing interest rate r_b^* in the interbank market at date 1. Further,*

(i) There exists some n^* such that, in equilibrium, banks 1 through n^*-1 issue $\hat{d}_i(r_b^*)$ digital claims, banks n^*+1 through $N+1$ issue $d_i^\tau(r_b^*)$ digital claims, and $d_{n^*}^* = \frac{(N+1)D}{\lambda} - \sum_{i \neq n^*} d_i^*$.

(ii) At date 1, banks in more productive zones borrow reserves from banks in less productive zones.

(iii) At date 4, there are net ex-post payment flows from more productive banks to less productive banks.

In our model, the need for intermediate liquidity to satisfy merchants' demands at time 3 restricts the amount of lending a bank can engage in at date 0. The interbank market at date 1 plays a critical role in helping productive banks meet their intermediate liquidity needs, and acts as a vehicle using which resources can be transferred from less productive to more productive zones. This allows investment to increase in the more productive zones, relative to the less productive ones. As a result, at date 4, there are settlement transfers from the more productive to the less productive zones.

4 FinTech and Innovations in the Payment System

We view the banking system as providing a means to distribute resources to productive investment possibilities. We define an equilibrium in our model as achieving a higher productive efficiency than another equilibrium if it leads to higher real output. We note that in our partial equilibrium environment, we cannot conduct a full welfare analysis. Instead, we will use the notion of productive efficiency to evaluate some of the innovations and technological changes that have been proposed to modernize the financial system. There are broadly two classes of innovations to the payment system that are relevant to us: Changes that affect consumers by facilitating payments, and those that affect institutions such as interbank settlement systems. Examples of the former include crypto-currencies and phone money, both of which typically operate outside the traditional payment system.[10] Examples of the latter include permissioned ledgers (usually based on the premise of the BlockChain).

There are three main frictions in this model which impede productive efficiency. First, the cost of transferring funds through the payment system (τ) acts like a tax on the most productive banks and therefore reduces their investment. In addition, this tax induces a coordination friction so that strategic banks mimic other banks to avoid incurring the cost of transferring payments. This can also lead to lower investment and output. Second, the interbank rate (r_b) leads to a friction because the most productive banks have to share some of the surplus from their productive investment with less productive banks, reducing the Finally, merchants' demand for interim liquidity (as given by the λ parameter) introduces a friction because it generates a short-term demand for cash. Absent this liquidity shock, in our deterministic model, banks could simply promise to pay all merchants out of final proceeds of production. The magnitude and interplay of these frictions in the attendant equilibria will be affected by technological changes, or "FinTech."

While we are not sufficiently clairvoyant to predict the specific technological innovations that will succeed and be pervasively adopted, we can suggest ways in which innovations might affect the liquidity flows we have documented. In the context of our model, innovations in consumer payment systems will reduce λ—the interim demand for cash—and will increase α—the expenditure outside the customer's banking system. Similarly, those innovations that affect settlement systems will reduce τ, the cost of settling claims through the payment system.

We present results both for economies in which the banking system serves production zones that are sufficiently homogeneous (as exhibited in Proposition 1; here the banking system does not reallocate capital), as well as those that are sufficiently heterogeneous (as outlined in Proposition 2; here the banking system does reallocate capital across zones). The difference between these two types of equilibria highlight how the payment system and

[10] Well-known crypto-currencies include Bitcoin, Ethereum, and Ripple. Phone money includes mobile money such as Mpesa. More broadly, electronic money stored on cards such as the Octopus Card in Hong Kong allows payments to be made outside the traditional payment system.

interbank market affect the real activity or productive efficiency of the banking system.

4.1 Bank liquidity and innovation in consumer payments

Suppose consumers turn to non-cash or non-demand-deposit methods to settle their claims. Consider a cryptocurrency such as BitCoin. An agent who decides to make a percentage of his purchases in Bitcoin will withdraw money from the banking system and transfer it to another agent in return for BitCoin. The BitCoin seller will deposit the money they receive into their own bank account. After this initial transaction, the amount of fiat money in the system will be the same, but the ongoing demand for fiat money for transactions purposes will be lower: agents will make some purchases with BitCoin that previously were made with fiat money. This will, ceteris paribus, lead to a reduction in the interim liquidity demand for money, as capture by our λ parameter. A structural change in λ will affect both the interbank rate and productive efficiency.

First, consider the effect of a decrease in λ on the interbank rate. Recall that in homogeneous economies there is a range of interbank rates consistent with equilibrium. Denote $\underline{r} = \frac{(\mu_1+\delta N)f'(\frac{D}{\lambda})-(1+\tau\alpha)}{\lambda(1-\alpha)}$ and $\bar{r} = \frac{\mu_1 f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)}$. The equilibrium interbank rate lies in the interval $[\underline{r}, \bar{r}]$. By contrast, if the productivity across zones is sufficiently different (i.e., the economy is heterogeneous), there is a unique interbank rate. In both cases, we can determine how the equilibrium rate (or its range) and productive efficiency change with changes in liquidity demand. Let r^*_{bo} denote the equilibrium interbank interest rate at date 1 in the old equilibrium, before the change in λ, and r^*_{bn} the corresponding interbank interest rate in the new equilibrium, after the change in λ.

When zones are sufficiently homogeneous in productivity, a change in liquidity demand (λ) has an ambiguous effect on the equilibrium interbank rate. However, somewhat counterintuitively, a decrease in merchant demand for cash can lead to an increase in the interbank interest rate when zones are heterogeneous in productivity. In both cases, the total output of the economy increases, as the reduction in λ allows each bank increases its lending.

Proposition 3 *Suppose that there is a small decrease in the demand for intermediate liquidity by merchants, λ. Then, each bank i increases its issuance of digital claims d_i. Further,*

(i) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity, as in Proposition 2, then the unique market-clearing interbank rate may rise or fall; that is, r^*_{bn} can be either greater or smaller than r^*_{bo}.

(ii) If $\delta \leq \hat{\delta}$, so that zones are approximately homogeneous in productivity, as in Proposi-

tion 1, then the range of interbank interest rates supported in equilibrium can increase or decrease.

In both cases in Proposition 3, changes in the interbank interest rate occur because of three effects that result from a fall in λ. First, there is an increase in $\frac{D}{\lambda}$, which increases the investable funds of each bank. This leads to a direct increase in the supply of funds in the interbank market. Second, increased access to funds leads each bank to increase its lending to entrepreneurs, which leads to increased demand for funds in the interbank market. Third, the increased lending effect is larger for the more productive banks, as compared to the less productive ones. The more productive banks are borrowers in the interbank market, so, all else equal, aggregate demand for funds increases. Overall, the net supply of funds in the interbank market at a given interest rate may increase or decrease when λ falls. In numerical examples, we find that when δ is sufficiently higher than $\underline{\delta}$, the third effect is large, and the demand effects outweigh the supply effect. Hence, r_b^* increases as λ falls. Conversely, when δ is sufficiently close to (but still higher than) $\underline{\delta}$, the supply effect dominates, and r_b^* falls as λ falls.

As Proposition 3 states, in all cases the fall in λ increases the lending of each bank. It is then immediate that aggregate output increases.

Corollary 3.1 *Suppose that there is a small decrease in the demand for intermediate liquidity by merchants, λ. Then, productive efficiency increases.*

Another, somewhat older, FinTech innovation that has been widely adopted is the infrastructure around e-commerce. For example, PayPal allows consumers to purchase items without exposing either credit card or bank account details to the seller. This has permitted commerce between merchants and consumers that have no long-term relationship given that it does not require trust or verification beyond the knowledge that the buyer has adequate funds in their PayPal account. Further, new technologies such as the ubiquitous credit card reader Square and other safe online payment options mean that commerce can take place among a broader range of participants. As these technologies spread, in our model, it becomes easier for entrepreneurs to purchase inputs from other zones. We therefore interpret these innovations as corresponding to an increase in α, the proportion of inputs purchased from other zones.

Proposition 4 *Consider a small increase in α, the proportion of inputs purchased in foreign zones. Then,* *(i) If $\delta < \hat{\delta}$, so that zones are approximately homogeneous in productivity as in Proposition 1, each bank i continues to issue the same number of digital claims as before, $d_i = \frac{D}{\lambda}$. Further $\frac{\partial \underline{r}}{\partial \alpha} > 0$ and $\frac{\partial \bar{r}}{\partial \alpha} > 0$, so that the range of feasible interest rates shifts upward.*

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, then (a) a less productive bank i, that was issuing claims $\hat{d}_i^*$ in the old equilibrium, increases its issuance (b) a more productive bank j, that was issuing claims $d_j^{\tau *}$, decreases its issuance. The effect on the interbank interest rate r_b^* is ambiguous.

In homogeneous economies, an increase in α leads to an upward shift of the range of interest rates that may arise in the interbank market, even though it has no real effects. When α increases, digital claims issued by other banks are more likely to be cashed in at a given bank, which ceteris paribus increases the demand for ex ante liquidity. The interest rate increases to counter this demand. However, in heterogeneous economies, the increase in α actually has the effect of shifting production from the more productive to the less productive zones. All else equal (including the digital claims of all other banks), the increase in α reduces the cost to bank i of issuing digital claims, because it shifts the need for intermediate liquidity to other banks in the system. Yet, the more productive banks have to settle up at date 4 and incur the additional transaction cost τ, so the initial effect of increasing d_i operates asymmetrically across zones. The equilibrium effect is that less productive zones increase digital claims whereas more productive zones decrease digital claims. Thus, productive efficiency falls.

Corollary 4.1 *Consider a small increase in α, the proportion of inputs purchased in foreign zones. Then,*

(i) If $\delta < \hat{\delta}$, so that zones are approximately homogeneous in productivity as in Proposition 1, there is no effect on productive efficiency.

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, then productive efficiency falls.

The net effect of these two classes of innovations is ambiguous. Reducing the need for cash effectively relaxes the intermediate liquidity constraint that the bank faces, and therefore allows banks to increase the volume of "fountain pen money" that they produce. If there are investment opportunities, then this increases the productive efficiency of the banking system. By contrast, if banking systems become more integrated, then banks effectively lose control over their liquidity management and this can reduce the productive efficiency of the economy. A bank with very productive investment possibilities would like to disavow digital claims issued by banks with less productive investment possibilities, but is forced to honor them. This illustrates the relationship between the payments function of banks and the lending or investment function of banks.

4.2 Bank liquidity and innovation in the interbank payment system

Various payment processors have discussed changing the interbank settlement process. For example, Ripple offers a back office system to facilitate interbank payments.[11] In addition various settlement systems based on permissioned ledgers have been discussed.[12] Such changes will lead to a decrease in τ, the cost of ex post settlement. Our model implies that this will tend to increase the interbank rate.

Proposition 5 *Suppose that the ex post settlement cost τ decreases by a small amount. Then,*

(i) If $\delta < \hat{\delta}$, so that zones are approximately homogeneous in productivity as in Proposition 1, each bank i continues to issue the same quantity of digital claims as before, $d_i^* = \frac{D}{\lambda}$. Further, $\frac{\partial \bar{r}}{\partial \tau} = 0$ and $\frac{\partial \underline{r}}{\partial \tau} > 0$, so the range of feasible interbank interest rates decreases.

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, $r_b^{n^*} < r_b^{o*}$, then (a) a less productive bank i, that was issuing claims $\hat{d}_i^*$ in the old equilibrium, decreases its issuance (b) a more productive bank j, that was issuing claims $d_j^{\tau *}$, increases its issuance. The interbank interest rate, r_b^*, increases.

Lower transfer costs at date 4 strictly increase the interbank rate at date 1 when productivity is heterogeneous, and increase the range it can lie in when productivity is homogeneous. In other words, the cheaper is ex post liquidity, the more expensive is interim liquidity. When productivity is sufficiently heterogeneous across different zones, a decrease in τ leads to production being transferred from less productive to more productive zones. The opportunity cost for less productive banks increases sufficiently so that they transfer resources to the more efficient banks. Thus, more productive banks increase their issuance of digital claims. The immediate effect is an increase in the demand for funds in the interbank market, and hence in the interest rate. This, in turn, induces the less productive banks to invest less locally and lend money on that market. Therefore, productive efficiency increases.

[11] See `https://ripple.com/` .

[12] For example, see The Clearing House's proposed RTP rails or the SWIFT blockchain initiative.

Corollary 5.1 *Suppose that the ex post settlement cost τ decreases by a small amount. Then,*

(i) If $\delta < \hat{\delta}$, so that zones are approximately homogeneous in productivity as in Proposition 1, productive efficiency is unchanged.

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, then productive efficiency incresases.

5 Conclusion

While the role of banks in the economy has received increased scrutiny since the financial crisis, the implications of their role of in the payment system has largely been ignored. In this paper we have tried to clarify the links between household cash deposits, demand deposits that include bank lending, the interbank market and the payment system. Even though banks can freely trade reserves in an interbank market, strategic considerations that arise from the payment system affect where credit is allocated, and thus the productive efficiency of the economy. In other words, the amount a bank lends is critically affected by the fact that it functions as a part of the payment system.

For convenience, we assume that only banks participate in the interbank market, and that all banks are price-takers in this market. In reality, the central bank takes an active interest in the market clearing price and monetary policy often leads to it intervening in the market. If many different prices are consistent with equilibria, the central bank could use the market-clearing price as a coordination mechanism. Alternatively, the central bank can affect the tradeoff faced by banks with different local investment opportunities, and therefore affect productive efficiency.

In our model, we also assume that banks are the only source of external finance for entrepreneurs. A large literature has documented that firms borrowing from particular banks experience adversity when the lending bank is hit by a negative liquidity shock. For example, Khwaja and Mian (2008) show that small firms in Pakistan experience an increase in financial distress when their lending bank has a negative liquidity event. In a similar vein, Shenoy and Williams (2017) show that the supply of trade credit from a supplier to a customer depends on the supplier's access to bank credit. Thus, while other important sources of external finance clearly exist, the role of banks is critical.

Finally, our results take as given and highlight the importance of the payment system function of banks. A standard intuition is that in the presence of an interbank market to reallocate resources, the marginal product of capital and hence investment will equalize across different production regions. As we have demonstrated, if banks are also responsible for ex post settlement in the payment system, this will not occur. Indeed, we have illustrated both cross-sectional differences in investment, and more importantly, the occurrence of aggregate differences in output from the ideal benchmark level.

Appendix: Proofs

Proof of Lemma 1

Consider π_i, the payoff of bank i, as shown in equation (6). The derivative of π_i with respect to d_i is:

$$\frac{\partial \pi_i}{\partial d_i} = \begin{cases} \mu_i f'(d_i) - 1 - r_b r\lambda(1-\alpha) & \text{if } d_i < \frac{1}{N}\sum_{j\neq i} d_j \\ \mu_i f'(d_i) - (1+\tau\alpha) - r_b\lambda(1-\alpha) & \text{if } d_i > \frac{1}{N}\sum_{j\neq i} d_j. \end{cases} \tag{13}$$

The second derivative is $\frac{\partial^2 \pi_i}{\partial d_i^2} = \mu_i f''(d_i) < 0$, so that π_i is strictly concave in d_i.

Now, consider the region $d_i < \frac{1}{N}\sum_{j\neq i} d_j$. Then, setting

$$\mu_i f'(d_i) - 1 - r_b\lambda(1-\alpha) \quad = \quad 0 \tag{14}$$

yields $f'(d_i) = \frac{1+r_b r\lambda(1-\alpha)}{\mu_i}$, or $d_i^* = h\left(\frac{1+r_b+\lambda(1-\alpha)}{\mu_i}\right) = \hat{d}_i$.

Similarly, over the region $d_i > \frac{1}{N}\sum_{j\neq i} d_j$, setting

$$\mu_i f'(d_i) - (1+\tau\alpha) - r_b\lambda(1-\alpha) \quad = \quad 0 \tag{15}$$

yields $f'(d_i) = \frac{1+r_b\lambda(1-\alpha)+\tau\alpha}{\mu_i}$, or $d_i^* = h\left(\frac{1+r_b\lambda(1-\alpha)+\tau\alpha}{\mu_i}\right) = d_i^\tau$. ■

Proof of Lemma 2

Suppose the statement of the Lemma is incorrect. That is, suppose that $\tau > 0$ and for some banks i, j where $j > i$ (and so $\mu_j > \mu_i$), we have $d_j^* < d_i^*$. Note that $j > i$ implies that $d_j^\tau > d_i^\tau$ and $\hat{d}_j > \hat{d}_i$. Further, given the bank's best response function, it must be that $d_k^* \in [d_k^\tau, \hat{d}_k]$ for each k. Therefore, $d_j^* < d_i^*$ implies that $d_i^* > d_i^\tau$ and $d_j^* < \hat{d}_j$.

There are now two possibilities for bank i. Either (a) $d_i^* = \hat{d}_i$, in which case, from the best response function, we know that $d_i^* \leq \frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_j^*$, or (b) $d_i^* < \hat{d}_i$, in which case $d_i^* = \frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_j^*$. In both cases, it must be that

$$d_i^* \quad \leq \quad \frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_j^*. \tag{16}$$

Similarly, there are two possibilities for bank j. Either (a) $d_j^* = d_j^\tau$, in which case, from the best response function, we know that $d_j^* \geq \frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_i^*$, or (b) $d_j^* > d_j^\tau$, in which case $d_j^* = \frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_i^*$. In both cases, $d_j^* \geq \frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_i^*$. Write this inequality as

$$\frac{1}{N}\sum_{k\neq i,j} d_k^* + \frac{1}{N}d_i^* \quad \leq \quad d_j^*. \tag{17}$$

Adding the corresponding sides of (16) and (17) and simplifying, we have

$$\left(1+\frac{1}{N}\right)\, d_i^* \quad \leq \quad \left(1+\frac{1}{N}\right)\, d_j^*, \tag{18}$$

or $d_i^* \leq d_j^*$, which directly contradicts the assumption that $d_j^* < d_i^*$. ■

Proof of Proposition 1

Recall that, from the best response function for bank i, in any equilibrium we have $d_i^* \in [d_i^\tau, \hat{d}_i]$. Further, $k_i^* = d_i^*$ because all inputs are provided at a price per unit of a dollar.

Suppose that $d_{N+1}^\tau < \hat{d}_1$. Consider any $d \in [d_{N+1}^\tau, \hat{d}_1]$. It is immediate that $d \in [d_i^\tau, \hat{d}_i]$ for each i, because $d_i^\tau \leq d_{N+1}^\tau$ and $\hat{d}_i \geq \hat{d}_1$. Therefore, if each bank $j \neq i$ sets $d_j^* = d$, from the best response function, it is a best response for bank i to set $d_i^* = d$. In other words, for each such d, there is an equilibrium in which each bank i sets $d_i^* = d$, or alternatively sets $k_i^* = d$.

For any bank i, if $d^* \in [d_i^\tau, \hat{d}]$, is an equilibrium, then the interest rate in the interbank market must be such that $\frac{D}{\lambda}$ is in this interval. In particular, $d_{N+1}^\tau \leq \frac{D}{\lambda}$ implies that

$$h\left(\frac{1+r_b\lambda(1-\alpha)+t\alpha}{(\underline{\mu}+N\delta)}\right) \quad \leq \quad \frac{D}{\lambda}, \tag{19}$$

or $r_b \geq \frac{(\mu_1+\delta N)f'(\frac{D}{\lambda})-(1+t\alpha)}{\lambda(1-\alpha)} \overset{d}{=} \underline{r}$. Similarly, $\hat{d}_1 \geq \frac{D}{\lambda}$ implies that

$$h\left(\frac{1+r_b\lambda(1-\alpha)}{\underline{\mu}}\right) \quad \leq \quad \frac{D}{\lambda}, \tag{20}$$

or $r_b \leq \frac{\underline{\mu}f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)} \overset{d}{=} \bar{r}$.

Finally, observe that $\underline{r} \leq \bar{r}$ if and only if $\delta \leq \frac{\tau\alpha}{Nf'(\frac{D}{\lambda})}$.

Now, suppose that $\delta \leq \frac{\tau\alpha}{Nf'(\frac{D}{\lambda})}$, and consider any $r \in [\underline{r}, \bar{r}]$. It follows that there is an equilibrium in which $r_b^* = r$, and each bank plays a best response by setting $d_i^* = \frac{D}{\lambda}$ given that all other banks set $d_j^* = \frac{D}{\lambda}$ for each j. If each bank issues the same amount of digital claims, the net payment flows are zero. ■

Proof of Lemma 3

Suppose that $\tau > 0$. From Lemma 1, it follows that in any equilibrium, the quantity of digital claims issued by bank i will lie weakly between $d_i^\tau(r_b)$ and $\hat{d}_i(r_b)$, where r_b is the interest rate in the interbank market. Now, consider the inequality $\hat{d}_i < d_{i+1}^\tau$, for $i = 1, \cdots, N$. This inequality can be written as:

$$h\Big(\frac{1+r_b\lambda(1-\alpha)}{\mu_i}\Big) \quad < \quad h\Big(\frac{1+r_b\lambda(1-\alpha)+\tau\alpha}{(\mu_i+\delta)}\Big). \tag{21}$$

As $h(\cdot)$ is a decreasing function, the inequality holds if $\frac{1+r_b\lambda(1-\alpha)}{\mu_i} > \frac{1+r_b\lambda(1-\alpha)+\tau\alpha}{(\mu_i+\delta)}$, or

$$\delta \;>\; \frac{\mu_i\tau\alpha}{1+r_b\lambda(1-\alpha)}. \tag{22}$$

Now, this inequality must hold for all interest rates r_b that are feasible in equilibrium. From Lemma 2, $d_1^* \leq d_j^*$ for all $j > 1$. As the market-clearing condition requires that the average quantity of digital claims across banks must equal $\frac{D}{\lambda}$, it must be that $d_1^* \leq \frac{D}{\lambda}$. Further, if there is any $i \in \{1, \cdots, N\}$ such that $d_i^* < d_{i+1}^*$, then it must be that $d_1^* < \frac{1}{N}\sum_{j=2}^{N+1} d_j^*$. Then, from Lemma 1, we have $d_1^* = \hat{d}_1$.

Now, in equilibrium, market-clearing requires that $\sum_{i=1}^{N+1} d_i^* = \frac{(N+1)D}{\lambda}$. As d_1^* is the lowest quantity of digital claims across all banks, it must be that $d_1^* \leq \frac{D}{\lambda}$. Substitute $d_1^* = \hat{d}_1$; then

$$\begin{aligned} h\Big(\frac{1+r_b\lambda(1-\alpha)}{\mu_1}\Big) &\;\leq\; \frac{D}{\lambda}, \\ \text{Or,}\quad 1+r_b\lambda(1-\alpha) &\;\geq\; \underline{\mu}f'(\frac{D}{\lambda}). \end{aligned} \tag{23}$$

Therefore, the smallest value of $[1+r_b\lambda(1-\alpha)]$ in any equilibrium is $\underline{\mu}f'(\frac{D}{\lambda})$. That is, $r_b^* \geq \frac{\underline{\mu}f'(\frac{D}{\lambda})}{\lambda(1-\alpha)}$. We show below that this condition implies $d_1^* < d_j^*$ for all $j \neq 1$, which further strengthens the weak inequality on r_b^* to a strict inequality.

Substitute the expression $\underline{\mu}f'(\frac{D}{\lambda})$ in for $1+r_b\lambda(1-\alpha)$ on the RHS of inequality (22). We obtain

$$\delta \;>\; \frac{\mu_i\tau\alpha}{\underline{\mu}f'(\frac{D}{\lambda})}. \tag{24}$$

Inequality (24) must hold for all $i = 1, \cdots, N$. The maximal value of the RHS is when $i = N$, and in this case $\mu_i = \underline{\mu} + N\delta$. Substituting this in on the RHS,

$$\begin{aligned} \delta &\;>\; \frac{(\underline{\mu}+N\delta)\tau\alpha}{\underline{\mu}f'(\frac{D}{\lambda})} \\ \text{Or,}\quad \delta &\;>\; \frac{\underline{\mu}\tau\alpha}{\underline{\mu}f'(\frac{D}{\lambda}) - N\tau\alpha} = \underline{\delta}. \end{aligned} \tag{25}$$

Note that $\underline{\delta} > 0$ whenever $\tau < \frac{\underline{\mu}f'(\frac{D}{\lambda})}{N\alpha}$. Notice, from assumption 1, $\frac{\underline{\mu}f'(\frac{D}{\lambda})}{N\alpha} > \frac{1}{N\alpha}$, and so the restriction on τ follows from Assumption 1 part (ii).

Therefore, $\delta > \underline{\delta}$ implies that $\hat{d}_i < d_{i+1}^\tau$ for each $i = 1, \cdots, N$. In turn, that implies $d_1^* < d_2^* < \cdots < d_{N+1}^*$. ∎

Proof of Proposition 2

If $\delta > \underline{\delta}$ and $\tau > 0$, then, Lemma 3 applies, so that, in any equilibrium $d_1^* < d_2^* \cdots < d_N^*$.

Fix a value of r_b. Ignore market-clearing in the interbank market for now, and let $d_i^*(r_b)$ denote the best response of bank i in a Nash equilibrium in lending, given r_b. Observe that Lemma 1 implies that in any such Nash equilibrium, $d_i^*(r_b) \in [d_i^\tau, \hat{d}_i]$. Further, $d_1^\tau < d_2^\tau < \cdots < d_{N+1}^\tau$, and similarly $\hat{d}_1 < \hat{d}_2 < \cdots < \hat{d}_{N+1}$.

Now, define the excess supply of funds in the interbank market to be

$$Z(r_b) \quad = \quad \frac{(N+1)D}{\lambda} - \sum_i d_i^*(r_b). \tag{26}$$

Consider $r_1 = \frac{\mu f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)}$. Then, $\hat{d}_1(r_1) = \frac{D}{\lambda}$, so that for all $i > 1$, it must be that $d_i^*(r_1) \geq \frac{D}{\lambda}$. Further, from Lemma 3, we know the equilibrium interest rate r_b^* exceeds r_1. Therefore, there must exist at least one bank i such that $d_i^*(r_1) > \frac{D}{\lambda}$. Therefore, $Z(r_1) < 0$.

Next, consider $r_2 = \frac{\mu_{N+1} f'(\frac{D}{\lambda})-(1+\tau\alpha)}{\lambda(1-\alpha)}$. Then, $d_{N+1}^\tau(r_2) = \frac{D}{\lambda}$, so that for all $i < N+1$, it must be that $d_i^*(r_2) \leq \frac{D}{\lambda}$. As $\delta < \underline{\delta}$, it must be that for at least one zone i this inequality is strict. Therefore, $Z(r_2) > 0$.

Now, observe that d_i^τ and $\hat{d}_i$ are continuous in r_b. Therefore, $d_i^*(\cdot)$ and hence $Z(\cdot)$ are continuous in r_b. Hence, there exists some r_b^* such that $Z(r_b) = 0$.

To see uniqueness, observe that $Z(\cdot)$ is strictly decreasing in r_b. Therefore, there can be at most one value of r_b such that $Z(r_b) = 0$.

Finally, the form of the best responses given r_b^* follows from Lemma 1. ■

Proof of Proposition 3

(i) Suppose productivity is sufficiently heterogeneous across zones, with $\delta > \underline{\delta}$. Then, the equilibrium in Proposition 2 obtains. We know that the equilibrium has the form, $i = 1, \ldots j$ issue $\hat{d}_i$, whereas $i = j+1, \ldots N+1$, they issue d_i^τ.

Notice,

$$\frac{\partial \hat{d}_i}{\partial \lambda} \quad = \quad h'(d_i^*)\frac{r_b(1-\alpha)}{\mu_i} < 0$$
$$\frac{\partial d_i^\tau}{\partial \lambda} \quad = \quad h'(d_i^*)\frac{r_b(1-\alpha)}{\mu_i} < 0$$

Further, $\frac{\partial \hat{d}_i}{\partial r_b} = \frac{\partial \hat{d}_i}{\partial \lambda}\frac{\lambda}{r_b}$, and $\frac{\partial d_i^\tau}{\partial r_b} = \frac{\partial d_i^\tau}{\partial \lambda}\frac{\lambda}{r_b}$.

In any equilibrium, the market clearing condition holds:

$$\sum_{i=1}^{N+1} d_i^* - \frac{(N+1)D}{\lambda} \quad = \quad 0 \tag{27}$$

Thus,

$$\sum_{i=1}^{N+1} \frac{\partial d_i^*}{\partial \lambda} + \sum_{i=1}^{N+1} \frac{\partial d_i^*}{\partial r_b}\frac{\partial r_b}{\partial \lambda} + \frac{(N+1)D}{\lambda^2} \quad = \quad 0 \tag{28}$$

$$\sum_{i=1}^{N+1} \frac{\partial d_i^*}{\partial \lambda}\left(1 + \frac{\partial r_b}{\partial \lambda}\frac{\lambda}{r_b}\right) + \frac{(N+1)D}{\lambda^2} \quad = \quad 0 \tag{29}$$

Clearly, in equilibrium, it must be that $\left(1 + \frac{\partial r_b}{\partial \lambda}\frac{\lambda}{r_b}\right) > 0$. However, $\frac{\partial r_b}{\partial \lambda}$ may be either positive or negative, depending on the relative size of $\sum_{i=1}^{N+1} \frac{\partial d_i^*}{\partial \lambda}$ (the overall demand effect) as compared to $\frac{(N+1)D}{\lambda^2}$ (the supply effect).

(ii) Suppose productivity is sufficiently homogeneous across zones, with $\delta \leq \hat{\delta}$. Then, the equilibrium in Proposition 1 obtains. Consider a small decrease in λ. As the symmetric equilibrium quantity issued is $d_i^* = \frac{D}{\lambda}$ for each i, it follows that the quantity of digital claims issued by each bank increases.

Now, recall that $\underline{r} = \frac{(\mu_1+\delta N)f'(\frac{D}{\lambda})-(1+\tau\alpha)}{\lambda(1-\alpha)}$ and $\bar{r} = \frac{\mu_1 f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)}$. Therefore,

$$\begin{aligned} \frac{\partial \bar{r}}{\partial \lambda} &= \frac{1}{\lambda^2(1-\alpha)^2}\Big[\lambda(1-\alpha)\mu_1 f''(\frac{D}{\lambda})(-\frac{D}{\lambda^2}) - \{1-\alpha)(\mu_1 f'(\frac{D}{\lambda}) - 1\}\Big] \\ &= \frac{1}{\lambda^2(1-\alpha)}\Big[\mu_1\Big(-\frac{D}{\lambda}f''(\frac{D}{\lambda}) - f'(\frac{D}{\lambda})\Big) + 1\Big]. \end{aligned} \tag{30}$$

As $f(\cdot)$ is strictly concave, $f''(\frac{D}{\lambda}) < 0$, so that the term $-\frac{D}{\lambda}f''(\frac{D}{\lambda}) - f'(\frac{D}{\lambda}) + 1$ may be positive or negative. As a result, the effect on $\bar{r}$ is ambiguous.

Similarly,

$$\begin{aligned} \frac{\partial \underline{r}}{\partial \lambda} &= \frac{1}{\lambda^2(1-\alpha)^2}\Big[\lambda(1-\alpha)\mu_N f''(\frac{D}{\lambda})(-\frac{D}{\lambda^2}) - \{1-\alpha)(\mu_N f'(\frac{D}{\lambda}) - (1+t\alpha)\}\Big] \\ &= \frac{1}{\lambda^2(1-\alpha)}\Big[\mu_N\Big(-\frac{D}{\lambda}f''(\frac{D}{\lambda}) - f'(\frac{D}{\lambda})\Big) + 1 + t\alpha\Big]. \end{aligned} \tag{31}$$

Again, the effect on $\underline{r}$ is ambiguous.

As both $\underline{r}$ and $\bar{r}$ may increase or decrease when λ changes, the range too may either increase or decrease.

■

Proof of Proposition 4

(i) Suppose productivity is sufficiently homogeneous across zones, with $\delta < \hat{\delta}$, so that the equilibrium in Proposition 1 obtains. Consider an increase in α. Then, it is immediate that $d_i^* = \frac{D}{\lambda}$ is unaffected. (Note that here we need $\delta < \hat{\delta}$, because $\hat{\delta}$ itself falls when α increases.) By inspection, $\bar{r}$ increases as α increases. Further,

$$\begin{aligned} \frac{\partial \underline{r}}{\partial \alpha} &= \frac{1}{\lambda^2(1-\alpha)^2}\Big[\lambda(1-\alpha)(-\tau) + \lambda\{\mu_{N+1}f'(\frac{D}{\lambda}) - (1+\tau\alpha)\}\Big] \\ &= \frac{1}{\lambda(1-\alpha)^2}\Big[\mu_{N+1}f'(\frac{D}{\lambda}) - (1+\tau)\Big]. \end{aligned} \tag{32}$$

Given Assumption 1 (ii), it follows that the last expression is strictly positive. Therefore, $\frac{\partial \underline{r}}{\partial \alpha} > 0$.

(ii) Suppose productivity is sufficiently heterogeneous across zones, with $\delta > \underline{\delta}$, so that the equilibrium in Proposition 2 obtains. Consider a small increase in α. At the old equilibrium value, $\hat{d}_i(r_b^*)$ increases for each i. From Proposition 2, the banks offering $\hat{d}_i$ in the old equilibrium were the banks in less productive zones.

Consider the effect on d_i^τ. Depending on whether $\tau > r_b^*\lambda$ or $\tau < r_b^*\lambda$, keeping r_b^* fixed, d_i^τ may increase or decrease. Now, aggregate investment in the economy remains at $\frac{(N+1)D}{\lambda}$. Banks that were at $\hat{d}_i$ have increased their investment. Therefore, the most productive banks, who were at d_i^τ, must reduce their total investment.

As investments of all banks are changing, the net effect on r_b^* is ambiguous, and depends on the amount of the change and the respective marginal productivities of banks.

■

Proof of Proposition 5

(i) Suppose productivity is sufficiently homogeneous across zones, with $\delta \leq \hat{\delta}$. Consider a small decrease in τ to τ'. It must remain that $\delta < \frac{\tau'\alpha}{Nf'(\frac{D}{\lambda})} = \hat{\delta}$. Therefore, the equilibria exhibited in Proposition 1 continue to obtain. It is immediate that we still have $d_1^* = d_2^* = \cdots = d_{N+1}^* = \frac{D}{\lambda}$. Further, by inspection, $\underline{r}$ rises and $\bar{r}$ is unaffected.

(ii) Suppose productivity is sufficiently heterogeneous across zones, with $\delta > \underline{\delta}$, so that the equilibrium in Proposition 2 obtains. Consider a small decrease in τ. Such a decrease leads to an increase in $d^\tau(r_b^*)$ at a given value of r_b^*. Therefore, there is now excess demand in the interbank market; i.e., $Z(r_b^*) < 0$. To clear the market, r_b^* must increase. Therefore, $\hat{d}(r_b)$ decreases from the value in the previous equilibrium, so that, in the new equilibrium, the investment by banks 1 through n^* falls. Further, it must be that $r_b^{n*} > r_b^{o*}$.

■

References

[1] Acharya, Viral, Denis Gromb and Tanju Yoralmazer (2012) "Imperfect Competition in the Interbank Market for Liquidity as a Rationale for Central Banking" *, American Economic Journal: Macroeconomics*, 4(2): 184–217.

[2] Ashcraft, Adam, James McAndrews and David Skeie (2009), "Precautionary Reserves and the Interbank Market," Federal Reserve Bank of New York Staff Reports.

[3] Basel Committee on Banking Supervision (2008), "Principles for Sound Liquidity Risk Management and Supervision," Bank for International Settlements report.

[4] Bianchi, Javier and Saki Bigio (2014), "Banks, Liquidity Management and Monetary Policy" Federal Reserve Bank of Minneapolis Staff Report #503.

[5] Bech, Morten L., and Rod Garratt (2003), "The Intraday Liquidity Management Game," *Journal of Economic Theory* 109: 198–219.

[6] Committee on Payment and Settlement Systems, (2008) "The Interdependencies of Payment and Settlement Systems," Bank for International Settlements report.

[7] Diamond, Douglas and Philip Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *The Journal of Political Economy*, 91(3): 401–419.

[8] Donaldson, Jason, Giorgia Piacentino and Anjan Thakor (2016) "Warehouse Banking," forthcoming, *Journal of Financial Economics.*

[9] Edgeworth, F. (1888), "A Mathematical Theory of Banking," *Journal of the Royal Statistical Society* 51: 113–127.

[10] Freixas, Xavier and Bruno Parigi (1998), "Contagion and efficiency in Gross and Net Interbank Payment Systems" *Journal of Financial Intermediation* 7: 3–31.

[11] Freixas, Xavier, Bruno Parigi, and Jean-Charles Rochet (2000) , "Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank," *Journal of Money, Credit and Banking* 32(3): 611–638.

[12] Gorton, Gary and George Pennacchi (1990), "Financial Intermediaries and Liquidity Creation," *The Journal of Finance,* 45(1): 49–71.

[13] Gu, Chao, Fabrizio Mattesini, Cyril Monnet and Randall Wright (2013), "Banking: A New Monetarist Approach" *Review of Economic Studies* 80: 636–662.

[14] Kahn, Charles M. , James McAndrews, and William Roberds (2003), "Settlement Risk under Gross and Net Settlement," *Journal of Money, Credit, and Banking* 35(4): 591–608.

[15] Kahn, Charles M. and William Roberds (1998), "Payment System Settlement and Bank Incentives," *Review of Financial Studies* 11(4): 845–870.

[16] Kahn, Charles M. and William Roberds (2009a) "An Introduction to Payments Economics," *Journal of Financial Intermediation* 18(1): 1–23.

[17] Kahn, Charles M. and William Roberds (2009b)"Payments Settlement: Tiering in Private and Public Systems", *Journal of Money, Credit, and Banking* 41(5): 855–884.

[18] Khwaja, Asim and Atif Mian (2008), "Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market," *American Economic Review* 98(4): 1413–1442.

[19] Kohn, Meir (1999), "Early Deposit Banking," Working Paper, Dartmouth University.

[20] McAndrews, J. and S. Potter (2002) "Liquidity Effects of the Events of September 11, 2001" *Federal Reserve Bank of New York Economic Policy Review* 8(2).

[21] Norman, Ben, Rachel Shaw and George Speight, (2011) "The History of Interbank Settlement Arrangements: Exploring Central Banks' Role in the Payment System," Working Paper, Bank of England.

[22] Plantin, Guillaume (2015) "Shadow Banking and Bank Capital Regulation" *Review of Financial Studies* 28(1): 146–175.

[23] Rochet, Jean-Charles and Jean Tirole (2006), "Two-sided Markets: A Progress Report" *RAND Journal of Economics* 37(3): 645–667.

[24] Shenoy, Jaideep and Ryan Williams (2017), "Trade credit and the joint effects of supplier and customer financial characteristics," *Journal of Financial Intermediation* 29: 68–80.

[25] Tobin, James (1963) "Commercial Banks as Creators of "Money,"" Cowles Foundation Discussion Paper No. 159.

[26] Yanelle, Marie-Odile (1997), "Banking Competition and Market Efficiency" *Review of Economic Studies*, 64(2): 215–239.