M | MICHIGAN ROSS

Working Paper

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

Christine A. Parlour
Finance Group
University of California, Berkerly

Uday Rajan
Stephen M. Ross School of Business
University of Michigan

Johan Walden
Finance Group
University of California, Berkerly

Ross School of Business Working Paper
Working Paper No. 1337
December 2016

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

Christine A. Parlour† Uday Rajan‡ Johan Walden§

December 31, 2016

Abstract

We consider the interaction between the roles of a bank as a facilitator of payments in the economy and as a lender. In our model, banks make loans by issuing digital claims to an entrepreneur, who then uses them to pay for inputs. Issuing digital claims has two effects on a bank's liquidity. First, some of these claims used as payment are cashed in before the project is over, necessitating transfers in the inter-bank market to meet these intermediate liquidity needs. Second, the lending bank must transfer reserves to the other banks when the project is done to settle its claims. Each of these transfers has a cost; the endogenous interest rate in the inter-bank market and a settlement cost for final transfers. These costs, in turn, are frictions that affect bank lending. Banks in our model are strategic. If productivity is similar, a high cost of final transfers leads to a coordination friction and multiple equilibria, with each bank trying to match the average number of digital claims issued by other banks. We consider the effects of financial innovations (i.e., FinTech) on the payments system and show that a reduction in the need for intermediate liquidity can lead to an increase in the inter-bank interest rate, because it also induces each bank to increase its lending. We also show that innovations may shift investments from more productive to less productive regions.

*We thank Charlie Kahn, Andy Winton and seminar participants at the Central Bank of Ireland Microstructure Workshop, Colorado, Illinois, Minnesota Corporate Finance Conference, Oxford, the SEC, 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.

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

Abstract

We consider the interaction between the roles of a bank as a facilitator of payments in the economy and as a lender. In our model, banks make loans by issuing digital claims to an entrepreneur, who then uses them to pay for inputs. Issuing digital claims has two effects on a bank's liquidity. First, some of these claims used as payment are cashed in before the project is over, necessitating transfers in the inter-bank market to meet these intermediate liquidity needs. Second, the lending bank must transfer reserves to the other banks when the project is done to settle its claims. Each of these transfers has a cost; the endogenous interest rate in the inter-bank market and a settlement cost for final transfers. These costs, in turn, are frictions that affect bank lending. Banks in our model are strategic. If productivity is similar, a high cost of final transfers leads to a coordination friction and multiple equilibria, with each bank trying to match the average number of digital claims issued by other banks. We consider the effects of financial innovations (i.e., FinTech) on the payments system and show that a reduction in the need for intermediate liquidity can lead to an increase in the inter-bank interest rate, because it also induces each bank to increase its lending. We also show that innovations may shift investments from more productive to less productive regions.

1 Introduction

What is bank liquidity? It is variously measured and described by banks' ability to borrow reserves in the interbank market, banks' ability to satisfy customer demands for deposits, or banks' ability to satisfy interbank claims. While these forms of liquidity are clearly related, they are exchanged in different markets, among different players, and affect banks' profits in different ways. The purpose of this paper is to micro-found bank liquidity flows in each of these markets and determine how their interaction affects banking activity, banking competition and more broadly, the macroeconomy. This will give us a foundation to analyze the effect of technological change on the various forms of bank liquidity.

The size of interbank flows is massive. In the United States, the Federal Funds market, in which banks trade reserves averages daily volumes around $330bn USD.[1] The size of the payments system, in which banks net digital claims used by their customers to make payments is even larger. For example, CHIPS, the largest private clearing house, processed average daily volumes in 2015 of $1.436 trillion.[2] These flows are not completely exogenous: through their ability to make loans and so issue demand deposits, banks have a measure of control over these magnitudes.

In the US, effectively three currencies exist in parallel and trade at par, albeit in different markets. First, central bank reserves that banks' exchange among themselves, second, physical currency and third, digital claims issued by banks as receipts for deposits. The market for central bank reserves allows banks to borrow reserves or lend them, and effectively allows banks to transfer deposits between themselves. Physical currency is typically provided by households and withdrawn 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.

In the normal course of economic transactions, banks transform liquidity and demand between each of these markets. For example, a bank that receives a dollar in cash deposits, may convert this into reserves which it can trade with other banks in the interbank market. Alternatively, banks can create digital claims (typically for more than a dollar) which they issue to other consumers. Once a consumer pays for a good, her digital claim ends up in the seller's checking account, which could be with another bank. This necessities a net transfer of liquidity between banks which is done through the payments system.

In our model $N+1$ banks, with access to potentially different quality investments, have

[1] See Morten, Martin, McAndrews (2008).

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

the same, fixed amount of household deposits of cash. Banks serve their local continuum of entrepreneurs, who may wish to buy inputs from merchants affiliated with another bank. First, households deposit money in the bank. Then, banks trade deposits in the interbank market. This determines a market clearing price or interbank rate. At the same time, banks contract with entrepreneurs. In addition to funneling deposits to entrepreneurs, banks issue "fountain pen money."[3] This means that the size of demand deposits or digital claims is typically larger than the amount of physical currency deposited at any bank. Entrepreneurs use these digital claims to purchase inputs for their technology. There are two consequences to this. First, merchants holding digital claims may cash then in at the bank for cash. This interim demand for liquidity constrains the amount of digital claims banks issue compared to their cash deposits. 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, the amount of inside money or digital claims issued by banks and thus the real level of production in the economy. We take as exogenous households' deposit amounts, merchants' demand for cash and the cost of transferring reserves through the payments system. Liquidity does not flow smoothly in our model because of three basic frictions. First, banks face a deadweight cost to transfer funds through the payments system; this reduces the willingness to issue digital claims (compared to the first-best level) but increases it when there are benefits to being a net payer. Indeed, a strictly positive cost acts like a tax on the most productive banks, and reduces their investment. The second friction is the merchant's intermediate liquidity need, which requires funds to be transferred in the inter-bank market in anticipation of this demand. In the interbank market, the equilibrium price at which banks lend deposits acts as a 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 incur the cost of transferring payments. This generates multiple equilibria – not necessarily in the amount of inside money that they issue. That banks are strategic in interbank payments has been documented by McAndrews and Potter (2002), where they found such differences in reaction immediately after the 2001 bombing.

We characterize two classes of equilibria that differ in the amount of digital claims (inside

[3]This phrase is due to Tobin (1963).

money) that banks issue: those that are symmetric, which occur if local conditions between banking systems are sufficiently similar, and those that are heterogeneous so that the banking system is a tool for transferring resources to more productive banking systems. The first result that we establish is that these two types of economies are quite different. Specifically, in economies in which the productivity of banking systems is sufficiently similar, there is a coordination friction that manifests itself in stickiness in the interbank rate. The logic is as follows: banks know that they face a cost if they are a net payer in the payments system and so rationally try to issue only as many claims as the average bank in the economy. Because of this, the interbank rate, which clears in reserves, can only be pinned down in a range. This illustrates the feedback effects between the different types of liquidity. By contrast, if the economy features sufficiently heterogeneous banking systems, the interbank rate is uniquely determined. In such economies, more productive banks are net borrowers in the interbank market and net payers through the payments system.

Once we have established the properties of these two types of economies, we turn to the effect of financial innovation (i.e., FinTech). There are two types of innovations on which we focus our attention: Those that affect consumers and those that affect back office banking operations. Suppose we move to a world in which there is substantial innovation in consumer payments. We view this as having two effects: first on the demand for cash and second on the degree of integration across banking systems. There are two parameters in our model that correspond to such changes. We establish that a decreased demand for cash will lead to banks issuing more digital claims (so an increase in the money multiplier) and a weak increase in the interbank rate or the rate at which banks trade reserves. Additionally, if banking systems become more integrated then the money multiplier across banks shrinks and investments shift from more productive to less productive regions.

If the back office operations become more efficient, specifically, if efficiency increases in the payments system, this will lead to increases in the interbank rate. Further, it will case the money multiplier in different banks to converge. Clearly, some of these comparative statics are counterintuitive if the interbank rate is interpreted as an equilibrium in demand deposits. Our results show that it is important to consider all the sources of bank liquidity to determine how changes in one aspect affect market clearing prices in another.

Modeling details aside, our analysis has three important elements. First, banks facilitate payments which generate large, predictable inter-bank 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 managements in the banking system, one of the earliest papers 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 use their framework to analyze the macroeconomic implications of banking behavior after the crisis. 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.

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 payments 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] Freixas, Parigi and Rochet (2000) present a model that is similar in structure to ours. In their paper, banks are part of a network because consumers face uncertainty about where they will consume. However, their focus is on contagion and risk in the payments system and how the central bank can alleviate it. Similarly, Freixas and Parigi (1998) considers the efficiency of real time or gross settlement in a world in which consumers withdraw funds from distant banks. By contrast, we consider the interplay between the payments system, the interbank market and inside money.

In a recent paper, Plantin (2015) presents a model in which bank insolvency is socially costly because agents use bank claims to settle transactions. Banks do not internalize the social cost of forgone transactions when they choose their leverage, and are thus subject to regulation. Making regulation more stringent can increase banks' use of the opaque shadow banking sector, which can improve overall welfare. 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.

The seminal work of Diamond and Dybvig (1983) considers the maturity transformation

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

role of banks. Banks are special because they uniquely have access to long term investment projects, however this exposes liquidity-shock prone investors or depositors to risk because early liquidation of long term projects is inefficient; this may lead to rational bank runs. Deposit insurance eliminates this risk. Our analysis differs from theirs in three fundamental ways. First, banks have no special role in investments as our entrepreneurs own observable machines and may seek funding from local money lenders; second in a Diamond Dybvig world, banks are exposed to the risk of the underlying investment, however if banks are viewed as part of the payment system, they bear no project risk per se. Finally, in the Diamond and Dybvig world, banks are a conduit from households to the real economy. In our framework, banks strategically determine how much credit to extend.

The microfoundations of payments systems and in particular the benefits of costs of gross versus net settlement are examined by Kahn and Roberds (2009b), Kahn and McAndrews (2003), Kahn and Roberds (1998), and Bech and Garratt (2003) 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 so do as a result of liquidity shocks that arise because of large value transfers. To avoid any possibility of insolvency, we assume that banks, perfectly anticipating their liquidity needs, borrow enough to cover liquidity demand. Thus, the transfer of net payments that we characterize In the context of our model, the ex post net transfer of reserves is equivalent to out ex ante larger borrowing on the part of the large bank. 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 is focuses on intermediaries' role in transferring value and providing a payments system.[5] Both 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,

[5] See Kohn (1999) and Norman, Shaw and Speight (2011) for a description of early banking and settlement systems.

the banks facilitated a complicated barter arrangement between differing pairs of traders. This arrangement mirrors a net settlement system.

Kahn and Roberds (2009) provide an introduction to the economics of payment and settlement systems in the modern economy. They distinguish between wholesale or large-value settlement between two banks and retail or small-value settlement between a bank and a household. 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.

2 Model

An economy comprises $N+1$ banks, each of which is indexed by i. Each bank 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. Banking systems are distinct from each other with each having their own local markets. This may be interpreted as either geographic differentiation or product differentiation along some unmodeled criterion. We therefore refer to banking system i as zone i.

Three types of claims are exchanged between agents: fiat money (we refer to this as "cash" and denote it by c), digital claims issued by commercial banks when they accept cash for deposits and make loans (these are "demand deposits", denoted by d), and reserves that banks can trade among themselves or convert back to cash (these are "reserves", denoted by z). Although not formally included in the model, there is implicitly a central bank at which bank reserves are held and which can convert reserves into cash. In the United States, and in our model, cash, digital claims, and central bank reserves are all denominated in the same units and are exchangeable at par from one form to another. Finally, there are real inputs (k) that are employed in the the entrepreneur's technology to produce the real output. In our base model, the price of one unit of real inputs is normalized to 1. Relative to the price of real inputs, each of c, d, z, k have a unit price of 1. We emphasize the strategic and real frictions that prevent liquidity from flowing between the markets.

Liquidity transformation 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. Specifically, each bank i receives cash deposits of D. We focus on the effect of productivity differences across banks on liquidity creation and investment. To do so, we keep the resource base of households fixed in each zone, so that each bank receives the same level of deposits. 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.

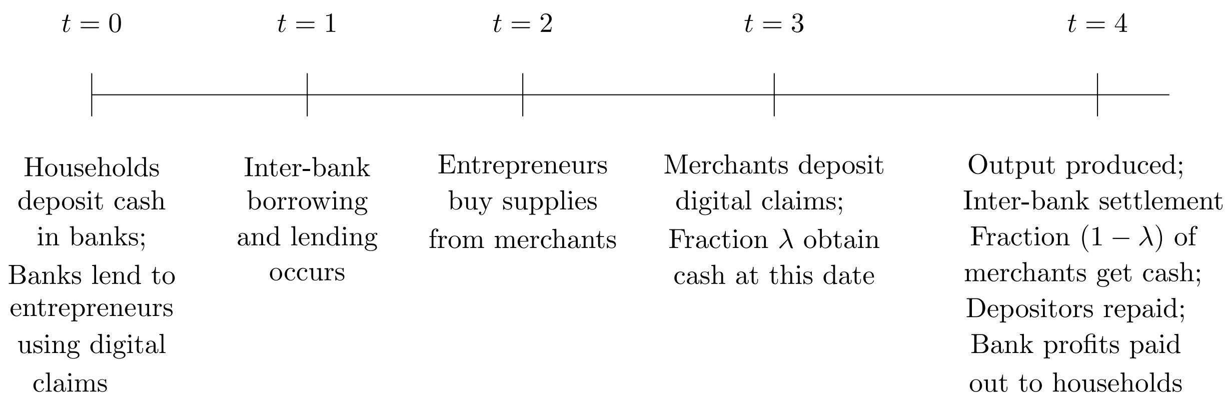

Figure 1: **Sequence of events.**

At $t = 1$, the inter-bank 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. The reason to trade reserves at this point is to accommodate banks' liquidity needs at date 3.

At $t = 2$, entrepreneurs use digital claims to purchase their inputs from merchants. Each entrepreneur has access to a strictly concave production technology, $f(k)$, where k is the quantity of real inputs they have purchased. The production technology is risky— for each entrepreneur in zone i, it succeeds with probability μ_i fails with probability $1 - \mu_i$. In the latter case, the machine produces nothing and all inputs are lost. The risk of project failure is idiosyncratic, with success probabilities being independent across entrepreneurs. Without loss of generality, we assume 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 assume that the difference in productivity between any two banking zones is constant, at $\mu_{i+1} - \mu_i = \delta \geq 0$.

Merchants are perfectly competitive, and the input price is normalized to 1 in expectation. We assume that the bank absorbs the risk of entrepreneur failure, rather than the merchants. As the bank funds a continuum of entrepreneurs, it faces no risk in the aggregate. Each merchant transacts with at most one entrepreneur in each zone. Nevertheless, merchants bear no risk, because the bank guarantees the payout on the digital claims it issues. Therefore, a digital claim for one unit of "money" procures a unit quantity of inputs. Let d_i denote the size of digital claims available to entrepreneur i, and k_i the quantity of inputs she purchases from merchants. We generally refer to d_i as the quantity of digital claims issued to entrepreneur i. In our base model, $k_i = d_i$. Later, in Section 3.1, we comment on the case in which $d_i > k_i$.

We introduce the payment system by assuming that entrepreneurs have to buy inputs in different locations. 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 are 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 digital claims they have received (from entrepreneurs in different zones) in their own bank. Note that a merchant's bank will often be different from the entrepreneur's bank. Collectively, each entrepreneur's purchase of inputs generates liquidity demand from the payments system at each merchant's bank. Specifically, a proportion $\lambda \in [0, 1]$ of merchants have a need for liquidity before the output is realized. These merchants arrive at their bank at date 3 and immediately convert their digital claims to cash. Each bank i therefore needs to either hold cash reserves (obtained at date 0 from households) or must borrow from other banks (at date 1) to meet this liquidity demand. A proportion $1 - \lambda$ of merchants 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 inter-bank settlements are made, and final remittances are made to merchants and households. When the output is produced, successful entrepreneurs deposit it back into their own bank, and the bank makes all payments. Inter-bank settlements are required whenever $\alpha > 0$. 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, which we denote by τ, for final settlement at date 4. This cost is an important exogenous parameter in our model. Through most of the paper, it may be thought of as the opportunity cost of funds that must be pre-deposited before bank i can participate in the payments system. Therefore, τ may fluctuate with the federal funds rate or a short-term repo rate. In Section 4.3, we consider the possibility that interchange fees earned by a bank that issues digital claims can reduce τ.

The costs borne by banks to transfer funds to other institutions are generally fees levied by independent entities, and are not determined by market-clearing. A few large systems are used in practice for inter-bank settlement. The fee structure and the economic costs of participating in each of these systems differs. However, for our purposes it is important to note that the cost of using such a system is positive, it increases in the volume of payments, and is determined not by market-clearing, but rather is set based on historical payment patterns. For example, the ACH system processes around 25 billion separate transactions a year. There are small administrative costs associated with accessing the network, and, in addition, originating a transfer incurs costs that range range between US $0.15 and 0.95 per transaction. The Fedwire system is used for same-day large payments. The fee structure includes discounts based on the average monthly volume; for small transactions (less than $14,000) the undiscounted fee is $0.79 per transfer.[6] 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, in addition to levying fees.[7]

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.

2.1 Liquidity Flows in the Economy

The liquidity flows for one zone are presented in Figure 2.

[6] https://www.frbservices.org/servicefees/fedwire_funds_services_2016.html

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

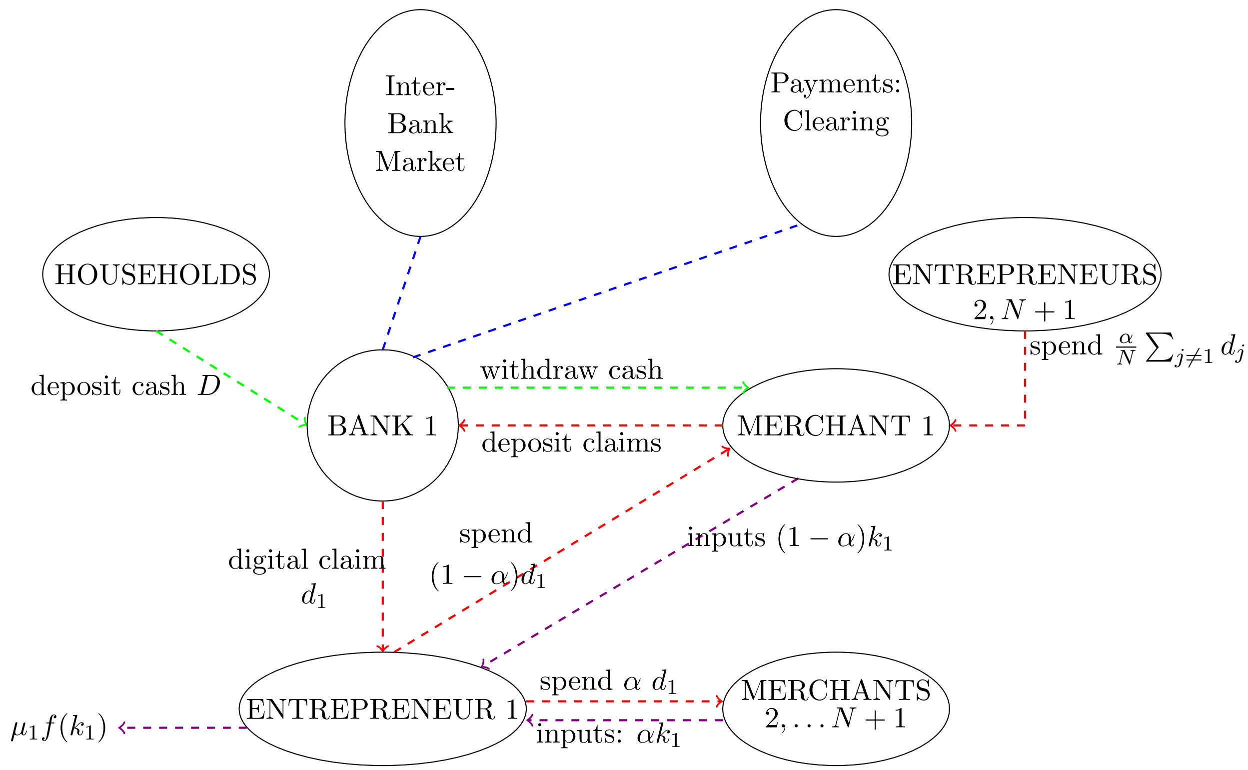

Figure 2: **Liquidity transformation in the economy.** The figure shows the flows of deposits, cash reserves, and digital claims for Zone 1. Cash flows are in green, digital claims are in red, real goods are in violet, and reserves are in blue.

Suppose that banks in the economy have issued digital claims $d_1, \ldots d_{N+1}$, and entrepreneurs have used them to buy inputs. The digital claims received by merchant i in exchange for real inputs are:

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

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 $j \neq i$.

Each merchant deposits these claims in their local bank at date 3. By this time, a proportion λ of the merchants in each zone have been hit by a liquidity shock. These merchants immediately convert their digital claims into cash. Thus, 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. Any deficit between λS_i and D must be borrowed in the inter-bank market. Conversely, a surplus can be lent out in the inter-bank market. Define $z_i = \lambda S_i - D$. The amount z_i may be positive or negative, and is the amount borrowed or lent by bank i in the inter-bank market at date 1. At date 4, inter-bank loans are also settled, and bank i has a cash flow of $-(1 + r_b)z_i$. We assume transfers across banks related to inter-bank loans are costless both at date 1 and at date 4.

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.

2.2 Bank Objective Function

Each bank seeks to maximize its profit. The bank's payoff function has several components. The first component is its share of the surplus from production. Suppose the bank issues digital claims in the amount d_i. 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 $\mu_i f(k_i) - k_i = \mu_i f(d_i) - d_i$. We normalize the reservation utility of the entrepreneur to zero; that is, to the utility obtained from zero units of cash. As the bank is large relative to the entrepreneur, we assume that the entrepreneur is held down to their reservation utility, and the bank captures the entire surplus from production. Finally, note that with a continuum of entrepreneurs receiving i.i.d. success or failure shocks, the bank faces no aggregate risk.

Note that cash is paid out for the inputs at different dates. An amount αd_i is paid to merchants in zone i who sell inputs to entrepreneur i. This amount is paid at two stages: $\lambda\alpha d_i$ is paid at date 3 to merchants who suffer the interim liquidity shock, and $(1-\lambda)\alpha d_i$ is paid at date 4 to merchants who are more patient. An amount $(1-\alpha)d_i$ must be transferred to other banks in the economy to pay for the inputs entrepreneur i buys in other zones $j \neq i$. As there is no discounting, the value at date 0 of future payments for inputs is equal to d_i.

The bank's profit from production is therefore $\mu_i f(d_i) - d_i$.

At date 3, bank i also pays cash on digital claims obtained by merchants in zone i when they sell to entrepreneurs from zones $j \neq i$. This outflow amounts to $\frac{\lambda(1-\alpha)}{N}\sum_{j\neq i} d_j$. At date 4, bank i further owes $\frac{(1-\lambda)(1-\alpha)}{N}\sum_{j\neq i} d_j$ to merchants in zone i for inputs provided for entrepreneurs $j \neq i$. In the inter-bank settlement flows at date 4, bank i is owed $\frac{1-\alpha}{N}\sum_{j\neq i} d_j$ from other banks. As the discount rate is zero, these flows net out to zero in present value, and do not enter the profit function.

The second component of the profit function is due to flows in the inter-bank 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 present value of these flows at date 0 is $r_b z_i$; that is, the interest cost or gain from borrowing or lending in the inter-bank market at date 1.

In addition, 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

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

The bank faces an interim liquidity constraint at date 3 — it must have enough cash on hand to meet merchant needs at that date. The demand for liquidity 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 inter-bank borrowing or lending, z_i. Note that we assume that inter-bank loans are immediately available as cash to the borrowing bank. The interim liquidity constraint at date 3 is therefore

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

It is easy to see that the cash-in-advance constraint (4) is binding. Suppose that it is slack, then 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,

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

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

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

While each bank takes the interbank rate r_b as given, it is an important equilibrium quantity. The interest rate in the inter-bank market is determined by market-clearing; 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

$$\sum_{i=1}^{N+1} d_i \;=\; \frac{(N+1)D}{\lambda}. \quad (7)$$

As we will show below, a bank's optimal quantity of digital claims, d_i, depends on r_b. Notice that 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.

2.3 Equilibrium

An equilibrium in this model is given by a Nash equilibrium in the banks' game given r_b, and an interest rate in the inter-bank loan market r_b that clears the market given banks' issuance of digital claims.

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

We impose two parameter restrictions. The first is a restriction on the marginal productivity of the least productive bank in autarky. Basically, the condition ensures that the bank would still invest all its resources even if there was no interbank market. The second restriction ensures that the cost of interbank transfers through the payment system, τ, is sufficiently low so that banks are willing to incur the costs.

Assumption 1 *The economy satisfies:*

(i) Sufficient Minimum Productivity: $\underline{\mu} f'(\frac{D}{\lambda}) - 1 > 0$

(ii) Bounded Transaction Costs: $\tau < \frac{1}{N\alpha}$*.*

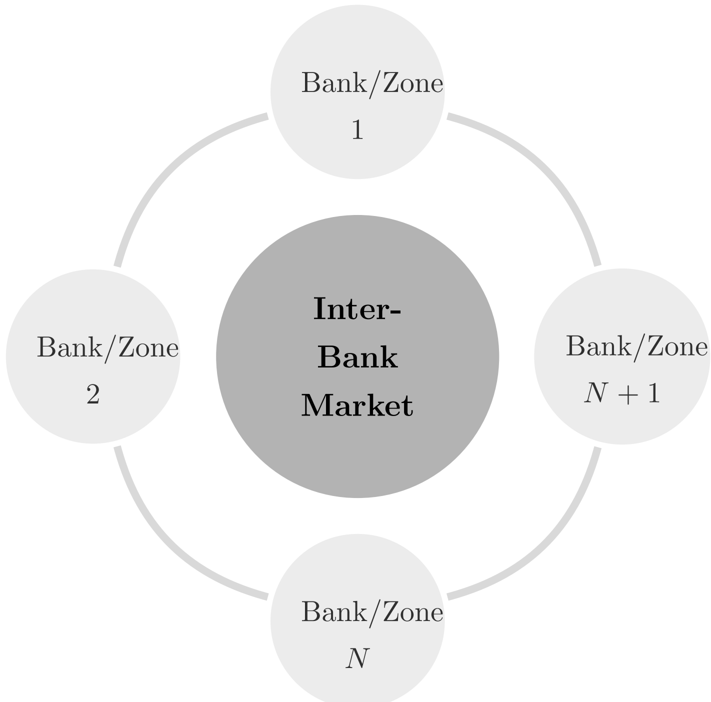

Figure 3: **Overall banking system.**

2.4 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 inter-bank market at date 1. 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 $h(x) \stackrel{\text{def}}{=} f'^{-1}(x)$. 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. In other words, higher marginal products are generated by lower input levels.

We define two input levels that 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 $\hat{d}_i < d_i^\tau$, whereas $\tau < 0$ 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 $d_i^\tau(r_b)$.

Suppose that $\tau > 0$. Fix the claims issued by banks $j \neq i$, and the interest rate in the inter-bank 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 compensate banks $j \neq i$ for the outflows they incur when merchants j cash in bank i's digital claims at bank j at date 3. 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 cost, in order to issue another digital claim, it must be that the productivity of the project is higher than the marginal cost of the input (fixed at 1) and the transfer cost.

The relationship between $\hat{d}_i$ and d_i^τ when $\tau > 0$ is illustrated in Figure 4 below.

Ceteris paribus, the cost τ, associated with netting out claims through the payments system at date 4, affects the best respose of a bank. As the proportion of foreign inputs (α) is the same across all zones, a bank is a net debtor at date 4 if the number of claims it has issued exceeds the average claims issued by the other banks. Thus, all banks take into account this average level of lending when they issue claims. Therefore, the best response of bank i depends on the average number of claims issued by banks $j \neq i$.

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*

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

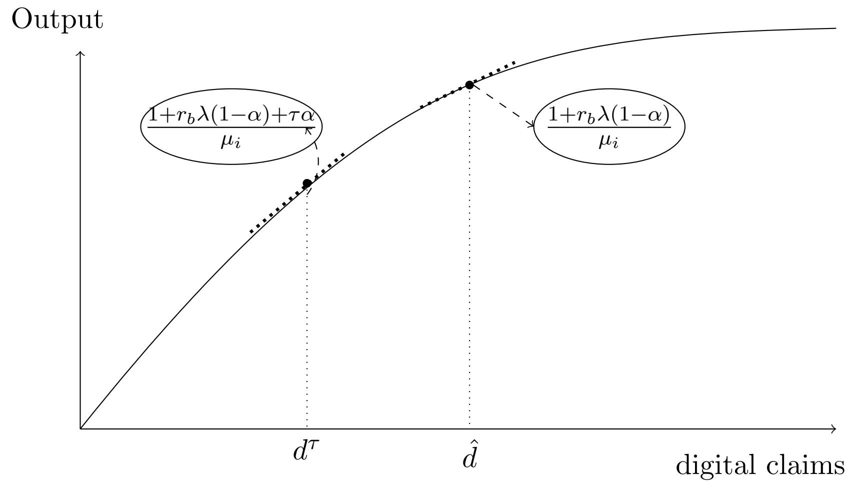

Figure 4: **Critical output thresholds when** $\tau > 0$. At d^τ, the marginal product includes the transfer cost, τ. When $\tau > 0$, this corresponds to a lower level of claims and inputs than at $\hat{d}$.

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, $\frac{1}{N}\sum_{j\neq i} d_j$.

Suppose that $\tau > 0$. Then, as mentioned above, $d_i^\tau < \hat{d}_i$. In this case, 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 inter-bank market at date 1, or in the final settlement at date 4. If, instead, it chose a higher level of digital claims d_i, it is then 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.

Conversely, suppose that $\tau < 0$. Then, $d_i^\tau > \hat{d}_i$. Then, there is no range at which bank i wants to match the average of the other banks. It either wants to be above that average, so that it can "pay" the negative transfer fee, or remain strictly below the average, to avail of the higher productivity (recall that the production technology $f(\cdot)$ is concave).

3 Positive Transfer Cost: $\tau > 0$

Throughout this section, we consider the case in which the cost of settlement transfers at date 4 is strictly positive; that is, $\tau > 0$. 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 *Suppose that $\tau > 0$. Then, 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 when $\tau > 0$. 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 the sense that the productivities across zones are not too different. 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 digitial claims than less productive zones. As each zone starts out with the same level of deposits, the inter-bank 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.

The simplest economy to consider is one in which there are no net flows in the interbank market or the net payment settlement market. This occurs if all banks serve entrepreneurs with similar productivities, and have the same level of deposits. In this world, the interbank rate is still defined, because it 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.

In these sufficiently symmetric economies, $d^* = \frac{D}{\lambda}$. Surprisingly, however, the coordination friction implies that the price that clears the interbank market is not necessarily unique, but will be in a range. Define

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

If $\delta \leq \hat{\delta}$, productivity is approximately homogeneous across zones.

Proposition 1 *Suppose that $\tau > 0$ and $\delta \leq \hat{\delta}$, 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 inter-bank 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, the gross payment flows in the economy total $\alpha(N+1)\frac{D}{\lambda}$, while the net payment flow between any two banks is zero.

Even though no banks have a net debt through the payments 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. The coordination friction that obtains for a fixed r_b, 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 even though each bank is utilizing its own deposits, this is not necessarily efficient. This is because banks have projects with differing success probabilities (or μ's). Intuitively, aggregate expected output is highest if the expected marginal products of capital are equalized across zones.

One of the important functions of capital markets is to reallocate funds from less productive to more productive areas. So, 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

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

First, we identify sufficient conditions under which the quantity of digital claims issued is strictly greater for banks with access to more productive projects.

Lemma 3 *Suppose that $\tau > 0$ and $\delta > \underline{\delta}$. Then, in any equilibrium,*

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

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

Under the conditions in Lemma 3, there is a unique equilibrium. While the aggregate investment in the economy is $\frac{(N+1)D}{\lambda}$ in each equilibrium, the distribution of investment across banks varies.

Proposition 2 *Suppose that $\tau > 0$ and $\delta > \underline{\delta}$, 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 inter-bank 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.

If there are differences in productivity, the more productive banks will borrow from the less productive banks in the ex ante interbank market. They do this because they will have to honor the claims of the merchants who cash in chits. They will also be net payers in the ex post market: they have issued more digital claims, and therefore will be net payers.

3.1 Discussion

In our model, all banks hold the same amount of deposits. Therefore, any differences in the volume of digital claims or inside money is effectively a difference in the money multiplier at the bank level. Specifically, the economy-wide multiplier must be $\frac{(N+1)D}{\lambda}$. However, the bank level multiplier may be higher or lower. In as much as local bank lending is related to local economic activity, such dispersion can have welfare effects.

We have presented two cases in which banks issue different levels of digital claims. The first is the case in which banks are ex ante symmetric, but the interest rate environment is sufficiently low so that there can be a net benefit to issuing more claims. In this case, even though banks have the same productivity, and the same level of deposits, some banks will issue more claims and hence have a higher mutiplier. The same is true in the case in which banks have different productivities, however in this case the more productive zones in the economy issue more claims and so have higher multipliers (for the same level of demand deposits).

So far, we have considered the case that the bank bears default risk. It is also interesting to consider the case in which each merchant bears risk. This could occur if entrepreneur failure is observed before the merchants cash in their claims. This is equivalent to the entrepreneur declaring bankruptcy. If the merchants bears default risk, then a risk neutral merchant values a given claim at μ_i. This will not affect the real side of the economy, but it will affect the nominal side. Specifically, the number of digital claims issued will increase. If a bank plans to induce a level of investment, k, then it will issue $\frac{k}{\mu_i}$ digital claims, knowing that merchants will discount them and only supply k units of input.

This increase in the inside money issued by banks will have no effect on the interbank rate. This is because if the merchant bears the risk then digital claims presented by entrepreneurs who subsequently go bankrupt cannot be redeemed at the bank. The merchant effectively loses the total amount of inputs that he supplied, and gets no compensation.

Finally, notice that the tendency of banks with riskier clienteles to issue more digital claims means that the more asymmetric the economy, the more distorted the relationship between claims and real investment. Indeed, any measure of demand deposits will not reflect real activity in the presence of default risk borne by the merchants.

4 FinTech and Innovations in the Payments System

Over the past few years, many innovations and technological changes have been proposed to modernize aspects of the financial system. Relevant to our are innovations to the payments system. These fall roughly into two classes: 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,[8] both of which typically operate outside the traditional payments system. Examples of the latter include permissioned ledgers (usually based on the premise of the BlockChain).

While we are not sufficiently clairvoyant to predict the innovations that will succeed and be pervasively adopted, we can suggest ways in which such innovations might affect the liquidity flows we have documented. In the context of our model, innovations in consumer payments systems will reduce λ, or the interim demand for cash, and will increase α the expenditure outside the customer's banking system. Similarly, those that affect settlement systems will reduce τ, the cost of settling claims through the payments system. We present these results for economies in which $\tau > 0$ and the banking system serves either entrepreneurs that are sufficiently symmetric, as exhibited in Proposition 1, or those that are sufficiently

[8]This could be mobile money such as Mpesa or electronic money on things such as the Octopus Card.

heterogeneous as outlined in Proposition 2.

4.1 Bank liquidity and innovation in consumer payments

What if consumers turn to non-traditional methods to settle their claims? This will, ceteris paribus, lead to a reduction in the ex ante demand for money, a variable that we capture with λ. A structural change in this parameter may affect all aspects of banking system liquidity. We focus our attention on inside money (digital claims) and the interbank rate. In the heterogeneous economies, there is a unique interbank rate however, in symmetric economies, there is a range of interbank rates. In spite of the multiplicity of interbank rates that are consistent with equilibrium, we can determine how the range changes with changes in underlying parameters. Denote $\underline{r} = \frac{(\mu_1+\delta N)f'(\frac{D}{\lambda})-(1+\tau\alpha)}{\lambda(1-\alpha)}$ and $\bar{r} = \frac{\mu_1 f'(\frac{D}{\lambda})-1}{\lambda(1-\alpha)}$. We examine how the ranges change in response to innovations that affect some of the parameters of the model.

Let r^*_{bo} denote the equilibrium inter-bank interest rate at date 1 in the old equilibrium, before the change in λ, and r^*_{bn} the corresponding inter-bank interest rate in the new equilibrium, after the change in λ. Somewhat counter-intuitively, a decrease in merchant demand for cash is accompanied by an increase in the inter-bank interest rate when zones are heterogeneous in productivity. This effect arises because λ also affects the banks' willingness to lend to entrepreneurs. A fall in λ leads to all banks increasing their lending (i.e., increasing the amount of inside money they create). This effect leads to an increased demand for money in the inter-bank market, counteracting the direct effect of lower λ. When zones are sufficiently homogenous on productivity, the net effect of these two forces is ambiguous.

Proposition 3 *Suppose that $\tau > 0$, and 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 $\delta \leq \hat{\delta}$, so that zones are approximately homogeneous in productivity as in Proposition 1, the effect on $\underline{r}$ and $\bar{r}$ is ambiguous, so the range of inter-bank interest rates supported in equilibrium may rise or fall.

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, then $r^*_{bn} > r^*_{bo}$. That is, the equilibrium inter-bank interest rate increases.

Next, consider changes in α, the proportion of inputs that the entrepreneur buys outside his banking system. New technologies such as the ubiquitous credit card reader "Square" and safe online payment options mean that commerce can take place among a broader range of participants. As these technologies spread, in our model, it is easier for entrepreneurs to purchase inputs from foreign zones. We therefore interpret these innovations as corresponding to increases in α, the proportion of inputs purchased from foreign zones.

Proposition 4 *Suppose that $\tau > 0$, and there is 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 $\frac{\partial \bar{r}}{\partial \alpha} > 0$ and $\frac{\partial \bar{r}}{\partial \alpha} > 0$, so that the range of feasible inter-bank interest rates shifts to the right.

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, the effect on the inter-bank interest rate r_b^* is ambiguous. The aggregate number of digital claims issued in the economy remains the same. However, (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.

That is, in homogeneous economies, an increase in α leads to an upward shift of the range of interest rates that may arise in the inter-bank market, even though it has no real effects. When α increases, digital claims issued by other banks are likely to be cashed in at any one bank which ceteris paribus increases the demand for ex ante liquidity, so the interest rate increases to counter this demand.

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 digitial claims of all other banks), the increase in α reduces the cost to bank i of issuing digital claims, because it shift the need for intermediate liquidity to other banks in the system. However, 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 and more productive zones decrease digital claims.

4.2 Bank Liquidity and Innovation in the Payments System

Various payment processors have discussed changing the interbank settlement process. For example, Ripple offers a back office system to facilitate interbank payments, in addition various settlement systems based on permissoned ledgers have been discussed.[9] Such changes will lead to a decrease in 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 is strictly positive and increases 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 inter-bank interest rates shrinks.

(ii) If $\delta > \underline{\delta}$, so that zones are sufficiently heterogeneous in productivity as in Proposition 2, $r_b^{n^*} < r_b^{o*}$, so that the inter-bank interest rate falls. The aggregate number of digital claims issued in the economy remains the same. However, (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.

Higher transfer costs at date 4 stricty reduce the interbank rate at date 1 when productivity is heterogeneous, and shrink the range it can lie in when productivity is homogeneous. In other words, the more expensive is ex post liquidity, the cheaper is interim liquidity. This negative correlation arises because even if liquidity is available at a low price in the interbank market, the more productive banks do not find it optimal to borrow and increase their lending: Doing so would expose them to transfer costs through the payments system.

When productivity is sufficiently heterogeneous across different zones, we again have the feature, as in Proposition 4, that the change implied by a fintech innovation leads to production being transferred from more productive to less productive zones. More productive banks cut back on their issuance of digital claims in order to save on the transactions cost. The immediate effect is a fall in the interest rate in the inter-bank market, which then induces the less productive banks to invest more locally rather than lend money on that market.

4.3 Asymmetric Transfer Fees Across Banks

So far, we have treated the cost of ex post transfers at date 4, τ, as being common across banks, and based on the opportunity cost of posting funds in advance of participating in the payments system. It is important to note that banks can also obtain some benefits from issuing a large quantity of digital claims (and thus being net payers at date 4). For example,

[9] `https://ripple.com/`

banks that issue credit or debit cards receive interchange fees from the merchant's bank, which serve to reduce the cost of being a net debtor.[10] Thus, in practice, the form in which digital claims are issued can affect the payments to be received by a bank at date 4.

Consider the case in which productivity is homogeneous across banks; that is, $\delta \leq \hat{\delta}$, where $\hat{\delta}$ is defined in equation (11). If all banks issue the same mix of digital claims, the interchange fees are a wash, in the sense that the fees paid by bank i will equal the fees received by it. As a result, the effective transfer cost τ will remain the same across banks, and the equilibrium in Proposition 1 will continue to apply.

A more interesting case arises if different banks choose different mixes of digital claims to issue. In this case, some banks earn higher interchange fees than others, leading to heterogeneous transfer costs at date 4. If, in addition, the federal funds rate is near zero, a bank i with high interchange fees effectively has a negative transfer cost. From the definitions of $\hat{d}_i$ and d_i^τ in equations (8)–(9), it follows that when $\tau < 0$, we have $d_i^\tau > \hat{d}_i$. If the productivity of the bank is sufficiently high, the bank then has an incentive to break the symmetric equilibrium exhibited in Proposition 1.

As a result, in equilibrium, banks that earn high interchange fees may issue a greater number of digital claims than other banks. Importantly, it need not be the case that these banks have the most productive investment opportunities (i.e., access to entrepreneurs with the highest values of μ). This provides another example in our model of a financial friction being converted into a real friction.

5 Conclusion

Since the financial crisis, the role of banks in the economy has received increased scrutiny. However, the role of the banks 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 in the economy. These strategic considerations are especially severe in low interest environments.

[10] Interchange fees are reimbursements offered to the issuing bank and are paid by the acquiring or merchants' bank. The actual fees depend on the type of card and how it is used. While fees can be as low as 5 basis points, for many standard credit products such as Visa signature preferred card, the fee is 2.1% of the transaction value. Thus, the cost of being a net debtor in the ex post market depends on the distribution of payment methods used by a bank's customers. A detailed description of Visa's 2016 interchange fees appears at `https://usa.visa.com/support/small-business/regulations-fees.html`

In this paper, we have assumed that banks are the only ones who participate in the interbank market and that all banks are price takers. In reality, the central bank takes an active interest in the market clearing price and intervenes. If many different prices are consistent with equilibria, the central bank could use the market clearing price as a coordination mechanism.

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 banking systems. As we have robustly demonstrated, if banks are also responsible for ex post settlement as in the case of the payment system this will not occur. Indeed, we have illustrated both cross sectional differences in investment and more importantly, aggregate differences in output that occur.

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 inter-bank 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 $r_b \geq \frac{(\mu_1 + \delta N) f'(\frac{D}{\lambda}) - (1+t\alpha)}{\lambda(1-\alpha)} \stackrel{d}{=} \underline{r}$. 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 $r_b \leq \frac{\underline{\mu} f'(\frac{D}{\lambda}) - 1}{\lambda(1-\alpha)} \stackrel{d}{=} \bar{r}$.

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

Now, suppose that $\delta \leq \frac{\tau\alpha}{N f'(\frac{D}{\lambda})}$, 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 while the gross flows are $\alpha(N+1)\frac{D}{\lambda}$ ■

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

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

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

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

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

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

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,

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

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 $d_1^* < d_2^* \cdots d_N^*$.

Given any value of r_b, let $d_i^*(r_b)$ denote the best response of bank i in the

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

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 $d_i^*(r_1) \geq \frac{D}{\lambda}$. From Lemma 3 there is some d_j^* for which the inequality is strict. 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 $d_i^*(r_2) \leq \frac{D}{\lambda}$. . Therefore, $Z(r_2) < 0$.

Now, $Z(\cdot)$ is continuous in r_b, so there exists some r_b^* such that $Z(r_b) = 0$.

To see uniqueness, observe that $Z(\cdot)$ is strictly decreasing in r. 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 homogeneous across zones, with $\delta \leq \hat{\delta}$. Then, the equilibrium in Proposition 1 obtains. Consider a small decrease in λ. Notice that an increase in λ results in an increase in $\hat{d}$, so in equilibrium, the number of digital claims is the same across all banks. As $d_i^* = \frac{D}{\lambda}$ 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,

$$\begin{aligned}
\frac{\partial \bar{r}}{\partial \lambda} &= \frac{1}{\lambda^2(1-\alpha)^2}\Big[\lambda(1-\alpha)\mu_1 f''(\frac{D}{\lambda})(-\frac{D}{\lambda^2}) - \{1-\alpha)(\mu_1 f'(\frac{D}{\lambda}) - 1\}\Big] \\
&= \frac{1}{\lambda^2(1-\alpha)}\Big[\mu_1\Big(- \frac{D}{\lambda} f''(\frac{D}{\lambda}) - f'(\frac{D}{\lambda})\Big) + 1\Big].
\end{aligned} \tag{27}$$

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,

$$\begin{aligned}
\frac{\partial \underline{r}}{\partial \lambda} &= \frac{1}{\lambda^2(1-\alpha)^2}\Big[\lambda(1-\alpha)\mu_N f''(\frac{D}{\lambda})(-\frac{D}{\lambda^2}) - \{1-\alpha)(\mu_N f'(\frac{D}{\lambda}) - (1+t\alpha)\}\Big] \\
&= \frac{1}{\lambda^2(1-\alpha)}\Big[\mu_N\Big(- \frac{D}{\lambda} f''(\frac{D}{\lambda}) - f'(\frac{D}{\lambda})\Big) + 1 + t\alpha\Big].
\end{aligned} \tag{28}$$

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

(ii) Suppose productivity is sufficiently heterogeneous across zones, with $\delta > \underline{\delta}$. Then, the equilibrium in Proposition 2 obtains. Consider a small decrease in λ. This decrease results in an increase in the aggregate investment in the economy, $\frac{ND}{\lambda}$. Further, both $\hat{d}_i$ and d_i^{τ} increase, so that at the old equilibrium interest rate, each bank increases the number of digital claims issued.

Therefore, at the old equilibrium value r_b^{o*}, there is an excess demand of funds in the interbank market; i.e., $Z(r_b^*) > 0$. To clear the market, r_b must rise; that is, $r_b^{n^*} > r_b^{o^*}$.

■

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,

$$\begin{aligned}\frac{\partial \underline{r}}{\partial \alpha} &= \frac{1}{\lambda^2(1-\alpha)^2}\Big[\lambda(1-\alpha)(-\tau) + \lambda\{\mu_{N+1}f'(\frac{D}{\lambda}) - (1+\tau\alpha)\}\Big] \\ &= \frac{1}{\lambda(1-\alpha)^2}\Big[\mu_{N+1}f'(\frac{D}{\lambda}) - (1+\tau)\Big]. \end{aligned} \tag{29}$$

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. Depending on whether $\tau > r_b^*\lambda$ or $\tau < r_b^*\lambda$, keeping r_b^* fixed, d_i^{τ} may increase or decrease. Now, consider banks that were at d_i^{τ}. Aggregate investment in the economy remains $\frac{ND}{\lambda}$. Banks that were at $\hat{d}_i$ have increased their investment. Therefore, the most productive banks 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 increase in τ to τ'. It must remain that $\delta < \frac{\tau'\alpha}{(N)f'(\frac{D}{\lambda})} = \hat{\delta}$. 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, $\underline{r}$ falls 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 increase in τ. Such an increase leads to a reduction in $d^\tau(r_b^*)$ at a given value of r_b^*. Therefore, there is now excess supply in the interbank market; i.e., $Z(r_b^*) > 0$. To clear the market, r_b^* must fall. Therefore, $\hat{d}(r_b)$ increases from the value in the previous equilibrium, so that, in the new equilibrium, the investment by banks 1 through n^* has increased. 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 (2015) "Warehouse Banking," Working Paper.

[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. , William Roberds, and James McAndrews (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, C., Roberds, W., (2009b)"Payments Settlement: Tiering in Private and Public Systems", *Journal of Money, Credit, and Banking* 41(5): 855–884.

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

[19] 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).

[20] 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.

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

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

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

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