Preliminary and Incomplete: Do not circulate.

On Cross-Border Payments and the Industrial Organization of Correspondent Banking*

Garth Baughman
Federal Reserve Board of Governors

Cathy Zhang
Purdue University

Ashley Wumian Zhao
Purdue University

September 15, 2022

Abstract

Despite advances in domestic payments arrangements, cross-border payments remain costly and slow. This paper builds a model of cross-border payments where market power in correspondent banking relationships reduces efficiency. We consider two policy experiments: the introduction by one country of an internationally held CBDC, and development of an internationally interoperable settlement system. Both policies can at least attenuate the inefficiencies in cross-border payments, but neither is automatically a complete solution nor are they without difficulties. For an international CBDC, we identify a political economy barrier: banks of the country potentially introducing the CBDC disproportionately suffer losses while benefits tend to accrue to foreign depositors, so a central bank concerned with its domestic banks' profitability may be unlikely to take such an action. Interoperability does not face the same barrier, as benefits accrue more symmetrically, but we argue that technical and other issues inherent to interoperability may be difficult to overcome.

*The views expressed are those of the authors and do not reflect those of the Federal Reserve System or the Board of Governors.

1 Introduction

Cross-border payments are a central component of the international monetary system. At the heart of cross-border transactions are correspondent banking relationships, in which one financial institution carries out transactions on behalf of another, often because it has no local presence.[1] The value of cross-border payments is estimated to increase from approximately $150 trillion in 2017 to over $250 trillion by 2027 (Bank of England, 2022). Moreover, while cross-border transactions represent approximately 20% of total transaction volumes in the payments industry, they generate 50% of its transaction-related revenues (McKinsey 2016).

Due to the vast scale and reach of cross-border transactions, there is an ongoing global effort across central banks, international financial institutions, and regulatory bodies to improve the cost, speed, and transparency of cross-border payments, which is notoriously more costly, slow, and opaque than domestic payments. However, despite the growing public interest and expansive literature on the broader nature of the international monetary system, there is a lack of formal modelling aimed at clarifying the underlying mechanisms that generate the inefficiencies in current cross-border payments. The objective of this paper is to develop a formal model of the central relationship at the heart of cross-border payments – correspondent banking – and use the model to demystify the key mechanisms influencing the efficiency of cross-border transactions. In addition, we analyse the effect of recent proposals aimed at improving efficiency such as the introduction of central bank digital currencies or policies promoting interoperability of payment systems.

Correspondent banking is the dominant arrangement for international payments where a bank in one country holds an account with a bank in a foreign country for the purposes of effecting international transfers. These relationships date to the beginnings of baking institutions themselves, perhaps as early as the original Venetian banks in the twelfth century. In its modern form, correspondent banking allows foreign banks access to a foreign country's payment systems without the expense of maintaining a local branch. Moreover, correspondent banking services can be a substantial profit center for large banks, and includes key operations such as international fund transfers, cash management services, check clearing, loans and letters of credit or foreign exchange services.

Whether due to switching costs, informational advantages, regulation, or other considerations, banks have market power over their customers. A key assumption in our model is the presence of market power in correspondent banking relationships. In our model, correspondent banking creates a vertical supply chain in the provision of international payments. Since suppliers (banks) at each level of the chain have market power, this arrangement results in double marginalization – the foreign bank charges a markup on the domestic bank which charges a further markup on

[1] Cross-border payments include both wholesale and retail payments. While wholesale payments are typically between financial institutions, retail payments are between individuals and businesses and include international remittances.

retail depositors. We provide a micro-founded model featuring such a banking supply chain, and consider its implications regarding recent policy suggestions for the improvement of cross-border payments.

One policy prescription regards the introduction of internationally-available central bank digital currencies. The supposition is that, were foreign banks to have access to direct claims on the local central bank, this would remove one chain in the supply chain, hence reducing costs. Our model supports this analysis, but with an important caveat. While such a policy would reduce the cost of cross-border payments, this would come largely from a loss in profit for local banks by cutting them out of the payment supply chain. A central bank with interest in the profits of its own banks would not support such a policy. This might offer some suggestion of why, despite years of discussion, only two central banks – China and the Bahamas – have, to date, actually introduced their own digital currencies. Moreover, in these cases, the technology is available only to domestic holders.

Interoperability is the second policy prescription analysed in this paper. The notion of interoperability refers to the interconnection of domestic payment systems. A number of technical, regulatory, and economic barriers exist that make it difficult to implement, and such considerations have occupied a great deal of efforts from officialdom. But, if the barriers are high, the promise is equally grand: that an international transfer could clear as quickly, easily, and, importantly, cheaply as a domestic one. We provide a simplified model of such interoperability and analyse its implications, which are mainly favorable as interoperability can directly overcome the double marginalization problem. This may not prove a panacea, however, as there are substantial technical and economic hurdles to the implementation of interoperability of the kind modeled in this paper.

2 Related Literature

Our model is related to monetary models with market power in the banking sector. Rocheteau, Wright, and Zhang (2018) and Lagos and Zhang (2022) feature banks with market power in over-the-counter markets for loans and settlement. Chiu et al. (2019), Keister and Sanches (2019), and Andolfatto (2021) focus on banks with market power in deposit markets and the implications on central bank digital currency.

This paper also contributes to the industrial organization literature on the banking system (Egan, Hortacsu, and Matvos 2017a; Egan, Lewellen, and Sunderam 2017b; Buchak, Matvos, Piskorski, and Seru 2018; Xiao 2018). While this work usually features a static industry equilibrium, we introduce a dynamic open-economy model where monopolistic banks charge mark-ups along the supply chain of payments.

3 Environment

There are two countries, home (H) and foreign (F). Each country is populated with four kinds of agents: buyers, sellers, entrepreneurs, and banks. Buyers and sellers are the households in the model who produce and consume goods and give rise to demand for deposits as a payment instrument. Buyers and sellers each have a population mass of one in each country. Entrepreneurs, have access to a productive investment, but no capital to invest. Each country has a single representative entrepreneur. Finally, each country has a single monopolist bank, which takes goods from households in exchange for deposits, and loans them to their local entrepreneur. Agents trade two perishable goods, a general good, X, and particular goods, q. The only asset is deposits issued by banks. Buyers and sellers may also engage in productive labor, H, which linearly produces the general good.

Time is discrete and continues forever, $t = 0, 1, \ldots$. Each period is divided into two sub-periods. In the second sub-period, there is a Centralized Market (CM) in which agents trade the general good, and banks redeem old deposits with interest and issue new ones, and entrepreneurs reap the fruits of old projects and invest in new projects. In the first sub-period, there is a Decentralized Market (DM) where buyers and sellers meet pairwise to trade particular goods. For simplicity, we assume buyers make a take-it-or-leave it offer to sellers. In the DM, buyers may meet a seller in their own country or abroad, with probabilities α and $1 - \alpha$, respectively. This is revealed in the previous CM. Since we have assumed symmetry across the two countries, the number of buyers attempting to contact a foreign seller is equal to the number of sellers there. So, we simply assume all sellers are contacted by some buyer, so the probability of matching in the DM is one.

In a DM meeting, the only payments technology is the transfer of bank deposits (there is no other good available to trade as the general good is perishable, and there is no other record keeping technology beyond bank deposits). Crucially, we assume that buyers and sellers can access only their local bank during the CM. This has two implications: First, buyers can only purchase deposits from their local bank, so their local bank has monopoly power in issuing deposits. Second, sellers can only accept payments drawn on their local bank, as these are the only deposits they can redeem. Combined, this implies that buyers who learn they will contact a seller abroad so require access to overseas deposits, can only acquire these deposits from their local bank. Hence, in order to offer their domestic buyers the ability to make international payments, banks engage in *correspondent banking* by first purchasing deposits from the other bank, and then offering them to their local customers.

Banks and entrepreneurs live only for one period, from one CM to the next, after which they are replaced with newly born clones (this eliminates complications due to banks or entrepreneurs accumulating savings over time, or developing reputations, etc.). Banks and entrepreneurs have linear preferences over the general good when they are old. Entrepreneurs have access to an investment technology that transforms k units of the general good invested in the CM when they are young into $f(k)$ units of the general good in the next CM when they are old. Entrepreneurs, however, are born without wealth, and can not produce the general good. Hence, they must borrow. But we assume entrepreneurs lack commitment, so can not directly take loans from households. Banks, instead, are assumed to have commitment and an enforcement technology over entrepreneurs. Hence, banks can take general good from households in exchange for deposits in one CM, give loan it to entrepreneurs, and reverse these transactions with interest in the following CM, consuming any profit. For simplicity, we assume banks make take-it-or-leave-it offers to entrepreneurs who, without any viable outside option are left with zero surplus. Therefore, one can consolidate banks with their captive entrepreneur, and we shall proceed as such, henceforth presuming banks have direct access to the investment technology. Alternately, perhaps due to regulation, we assume that banks must offer linear pricing to depositors, both their local households and the overseas bank, although we do allow price discrimination between the two.

Buyers and sellers live forever and discount the future with factor $\beta \in (0,1)$. Buyers have preferences over labor, H, consumption of the general good, X, and consumption of the particular good, q as follows:

$$U^B = U(X) - H + u(q).$$

Sellers have preferences over labor, H, consumption of the general good, X, and production of the particular good, q as follows:

$$U^S = U(X) - H - c(q).$$

Assume that preferences and technology – $U(\cdot)$, $u(\cdot)$, $c(\cdot)$, and $f(\cdot)$ – all have the usual properties, and, additionally, that $u(0) = c(0) = 0$. To avoid complications regarding the potential need for an alternative storage technology, assume that the investment technology has sufficient capacity. Specifically, for q^* solving $u'(q^*) = c'(q^*)$ assume that $\beta f'(q^*) > 1$. This completes the specification of the environment, and we now on to describe value functions and derive equilibrium.

4 Equilibrium

We will derive a stationary equilibrium recursively, beginning with buyers and sellers. For a buyer from country i who will, in the next period DM, meet a seller from country j, write $W_B^{i,j}(d)$ as the value of entering the CM with deposits which, cum interest, entitle the bearer to a claim on d general good. Note, as there is no bank default, and this is a real model, a buyer does not care whose claims they bring into the CM. This is not true in the DM, as sellers in country i can accept only deposits drawn on i. As we assume buyers know when entering the CM the nationality of the seller whom they will contact in the next DM, these buyers will only purchase deposits acceptable to that seller. As the countries are otherwise symmetric, we can, therefore, write $V_B^{i,j}(d)$ for the value of a buyer from country i who will meet a seller from country j entering the DM with deposits worth d units of general goods in the next CM (cum interest payments) as this is the relevant quantity determining how much special goods a buyer can purchase. It will be convenient to work with discounts instead of interest rates. So, write $\rho_B^{i,j}$ for the discount offered by bank i on transferable deposit claims on bank j. That is, in exchange for $\rho_B^{i,j}d$ units of the general good deposited in today's CM, bank i promises that bank j will repay d units of numeraire in tomorrow's CM (so the gross real interest rate is $R_B^{i,j} = 1/\rho_B^{i,j}$).

With this notation, we can write the buyer's problem in the CM:

$$W_B^{i,j}(d) = \max_{X,H,d^+} \left\{ U(X) - H + \beta V_B^{i,j}(d^+) \quad \text{s.t.} \quad X + \rho_B^{i,j} d^+ \leq d + H \right\}. \tag{1}$$

The value of a buyer entering the CM with claims to d general good is equal to the maximum attainable value of the sum of the utility of consumption of the general good, less production of the general good, plus the discounted value of continuing to the decentralized market with new deposit claims d^+ where this maximization is subject to the budget constraint that consumption of general goods, X, plus the current outlay for deposit claims, $\rho_B^{i,j}d^+$, must be less than the value of old claims plus any labor. Substituting out H from the budget constraint, one observes that the problem is separable into a linear value of old deposits, the surplus from consumption of the general good, and the value of buying new deposits and continuing to the next DM:

$$W_B^{i,j}(d) = d + \max_X \{U(X) - X\} + \max_{d^+} \left\{ -\rho_B^{i,j} d^+ + \beta V_B^{i,j}(d^+) \right\}.$$

A result of quasi-linear preferences, this separability implies that consumption and portfolio decisions are independent, and, further, that these decisions are independent of the wealth of the buyer. This independence greatly simplifies the aggregate environment, as buyers effectively "reset" each period, obviating dependence on individual histories and making any consequent distributions degenerate. The first order conditions of the buyer are

$$U'(X) = 1 \tag{2}$$

and

$$\frac{\partial}{\partial d^+} V_B^{i,j}\left(d^+\right) = \frac{\rho_B^{i,j}}{\beta}. \tag{3}$$

The envelope condition is

$$\frac{\partial}{\partial d} W_B^{i,j}(d) = 1.$$

This last result states that a buyer's CM value function can be written as

$$W_B^i(d) = d + W_B^i(0)$$

so is affine with unit slope, a dramatic simplification which will be used below. Sellers face a similar problem. But, as they make no purchases in the DM, they will purchase no deposits in the CM, instead simply redeeming any deposits they had from the previous DM, and working to achieve optimal consumption. While it will be immaterial (because we assume buyers make TIOLI offers to sellers) we note that sellers do not know in the CM which buyer they will contact in the following DM, so write $V_S^{i,j}$ for the value of a seller from i who meets a buyer from j in the DM.

$$W_S^i(d) = \max_{X,H} \left\{ U(X) - H + \beta \left[\alpha V_S^{i,i} + (1-\alpha) V_S^{i,\neg i} \right] \quad \text{s.t.} \quad X \leq d + H \right\}.$$

where we write $\neg i$ for the opposite country to i ($\neg i \equiv j | i \neq j$). Substituting out labor from the constraint, we find that, as with buyers, this problem is separable, and sellers have an affine CM value function:

$$W_S^i(d) = d + \max_X \{U(X) - X\} + \beta V_S^i.$$

Turning to the DM, suppose there is a meeting between a buyer from i, a seller from j, and the buyer has d units of deposits, cum interest, that the seller can accept (because the are drawn on bank j). In such a meeting, suppose the equilibrium quantity of particular goods produced for the buyer by the seller is $q_{i,j}(d)$ and the amount of deposits transferred is $e_{i,j}(d)$. The budget constraint for the buyer is that the transfer can not exceed their available balances: $e_{i,j}(d) \leq d$. The value function for a buyer is

$$V_B^{i,j}(d) = u(q_{i,j}(d)) + \hat{W}_B^i(d - e_{i,j}(d))$$

where we have written $\hat{W}_B^i(d) = \alpha W_B^{i,i}(d) + (1-\alpha) W_B^{i,\neg i}(d)$ for the expected value to a buyer of entering the next CM, where this expectation is taken over the matching shock. The value function for a seller is

$$V_S^{i,j} = -c(q_{i,j}(d)) + W_S^i(e_{i,j}(d)).$$

We can now solve for the equilibrium trade pair $(q_{i,j}(d), e_{i,j}(d))$. The buyer makes a take-it-or-leave-it offer to the seller. The seller's outside option is to produce nothing, and continue to the CM with no balances. Hence, their participation constraint given a trade pair (q, e) is

$$-c(q) + W_S^i(e) \geq -c(0) + W_S^i(0)$$

Because $W_S^i(d) = d + W_S^i(0)$ and $c(0) = 0$, this participation constraint simplifies to

$$e \geq c(q).$$

The buyer's problem when proposing a trade (q, e) is

$$\max_{q,e} \left\{ u(q) + \hat{W}_B^i(d - e) \quad \text{s.t.} \quad d \geq e \quad \text{and} \quad e \geq c(q) \right\}.$$

Given that $\hat{W}_B^i$ is affine, we can re-write this as

$$\max_{q,e} \left\{ u(q) - e + \hat{W}_B^i(d) \quad \text{s.t.} \quad d \geq e \quad \text{and} \quad e \geq c(q) \right\}.$$

Whereby the solution is immediate: If it is affordable, the buyer chooses the efficient quantity and just compensates the seller. If the efficient quantity is unaffordable, the buyer transfers all of his deposits, demanding a quantity that just satisfies the participation constraint. Writing q^* for the efficient quantity solving $u'(q^*) = c'(q^*)$, we have that the equilibrium trade satisfies

$$q_{i,j}(d) = \min\{q^*, c^{-1}(d)\}$$

and

$$e_{i,j}(d) = \min\{c(q^*), d\}.$$

With the DM trade solution in hand, we can now step back to characterize the demand for deposits in the CM. Given this solution, the envelope condition for the DM value function satisfies

$$\frac{\partial}{\partial d} V_B^{i,j}(d) = \max \left\{ \frac{u'(c^{-1}(d))}{c'(c^{-1}(d))}, 1 \right\}. \tag{4}$$

Substituting this into the first order condition for deposits, 3, one derives an expression for the demand for deposits, $D_B^{i,j}(\rho_B^{i,j})$:

$$D_B^{i,j}(\rho_B^{i,j}) = d \quad \text{solves} \quad \begin{cases} \frac{\rho_B^{i,j}}{\beta} = \frac{u'(c^{-1}(d))}{c'(c^{-1}(d))} & \text{if } \beta < \rho_B^{i,j} \\ d \in [c(q^*), \infty) & \text{otherwise.} \end{cases} \tag{5}$$

This completes the solution of the buyers' and sellers' problems.

We now turn to the characterization of the bank-entrepreneur problem. Recall, we assume banks and entrepreneurs live for one period, and only value consumption when old. Entrepreneurs have access to a productive technology which transforms k units of general good invested when young into $f(k)$ units of general good when old. Banks make take-it-or-leave-it offers to entrepreneurs, so the latter are left with no surplus. Hence, we can proceed as though banks have direct access to the productive technology. In order to fund their investment, banks issue deposits to households and the overseas bank. We assume that banks employ linear pricing, but can price discriminate between their domestic buyers and the other bank. Above, we derived buyers' demand for deposits, $D_B^{i,j}(\rho_B^{i,j})$. Write ρ_I^i for the discount offered on deposits by bank $\neg i$ to bank i (I for "interbank"), and write $D_I^i(\rho_I^i)$ for the demand for such deposits (to be derived below).

The problem of a bank is to choose a quantity of investment, k, the quantity of deposits it purchases from the other bank, d_I^i, and the discounts it sets on its deposits to maximize consumption when old subject to two constraints. The first is an inventory constraint – known as the *box constraint* in the finance literature (e.g. Huh and Infante (2021)) – and reflects the fact that intermediaries must be able to deliver the securities they promise. That is, while bank i can produce its own deposits on demand, it must purchase the other bank's deposits in order to deliver them. Hence, the box constraint requires that it purchase enough of those deposits to cover its promises:

$$d_I^i \geq D_B^{i,\neg i}(\rho_B^{i,\neg i}). \tag{6}$$

The second constraint is a simple budget constraint. As the bank is born poor, and can not work when young, its receipts from incoming deposits must exceed its outlays on capital and the other bank's deposits:

$$k + \rho_I^i d_I^i \leq \rho_I^{\neg i} D_I^{\neg i}(\rho_I^{\neg i}) + \rho_B^{i,i} D_B^{i,i}(\rho^{i,i}) + \rho_B^{i,i} D_B^{i,\neg i}(\rho^{i,\neg i}). \tag{7}$$

A bank's objective function is to maximize its consumption when old. When old the bank must pay out on its own deposits, both those to buyers, $D_B^{i,i}(\rho_B^{i,i})$, and those it sold to the other bank, $D_I^{\neg i}(D_I^{\neg i})$, and has receipts from the productive technology, $f(k)$, and any unsold deposits from the overseas bank, $d_I^i - D_B^{i,\neg i}(\rho_B^{i,\neg i})$.

These observations lead us to write the problem of bank i as

$$\max_{k, d_I^i, \rho_I^{\neg i}, \rho_B^{i,i}, \rho_B^{i,\neg i}} f(k) + d_I^i - D_I^{\neg i}(\rho_I^{\neg i}) - D_B^{i,i}(\rho_B^{i,i}) - D_B^{i,\neg i}(\rho_B^{i,\neg i})$$

subject to equations 6 and 7.

Given our assumption on the technology, $f(k)$, it will always be profitable to invest all receipts in the productive technology. Hence, the budget constraint holds with equality. Similarly, in any symmetric equilibrium with positive interest rates, it is easy to show that the box constraint holds exactly.[2] Substituting out the box constraint, and letting λ be the multiplier on the budget

[2] If not, then banks hold positive amounts of each other's deposits. For it to be profitable for bank i to supply those deposits, their investment project must earn more than the interest they're paying. But the interest bank i pays is equal to the interest bank $\neg i$ earns, which must then be less than the return $\neg i$ is earning on its investment. So, we must not be in a symmetric equilibrium. It remains for future work to consider the possibility of asymmetric equilibria.

constraint, one can form a Lagrangian

$$\mathcal{L} = f(k) - D_I^{\neg i}(\rho_I^{\neg i}) - D_B^{i,i}(\rho_B^{i,i}) \\ + \lambda\left[(\rho_B^{i,\neg i} - \rho_I^{i})D_B^{i,\neg i}(\rho_B^{i,\neg i}) + \rho_B^{i,i} D_B^{i,i}(\rho_B^{i,i}) + \rho_B^{\neg i} D_I^{\neg i}(\rho_B^{\neg i}) - k\right] \tag{8}$$

Writing

$$\epsilon_I^{\neg i} = \frac{\partial D_I^{\neg i}(\rho_I^{\neg i})}{\partial \rho_I^{\neg i}} \frac{\rho_I^{\neg i}}{D_I^{\neg i}(\rho_I^{\neg i})}$$

for the price elasticity of deposit demand of the overseas bank for local deposits, and similarly for the other demand curves, one derives familiar expressions for the equilibrium discounts on own deposits for local buyers and the overseas bank:

$$f'(k) = \frac{1}{\rho_I^{\neg i}} \frac{\epsilon_I^{\neg i}}{1 + \epsilon_I^{\neg i}}. \tag{9}$$

$$f'(k) = \frac{1}{\rho_B^{i,i}} \frac{\epsilon_B^{i,i}}{1 + \epsilon_B^{i,i}}. \tag{10}$$

The condition for pricing of overseas deposits offered to local buyers takes a different form, as proceeds, on the margin, are directed not to the investment technology, but instead in overseas deposits:

$$\rho_I^{i} = \rho_B^{i,\neg i} \frac{\epsilon^{i,\neg i}}{1 + \epsilon^{i,\neg i}}. \tag{11}$$

Given the other bank's demand and interbank rate, $D_I^{\neg i}(\rho_I^{\neg i})$ and ρ_I^i, equations 9-11 together with the budget constraint and the box constraint, equations 6 and 7, solve a bank's problem.

In principle, a full equilibrium solution could involve a complex fixed point problem in function space, as one might imagine this function for derived demand for interbank deposits would, itself, depend on the equilibrium interbank demand function. But, notice, there is a recursive structure to this problem. Equation 11 which determines $\rho_B^{i,\neg i}$ depends only on ρ_I^i and $D_B^{i,\neg i}(\rho_B^{i,\neg i})$, the latter of which is given explicitly by 5. Hence, using this value of $\rho_B^{i,\neg i}$ as a function of ρ_I^i, call it $P^i(\rho_I^i)$, and substituting into the box constraint directly gives one's own demand for interbank deposits *independently of the functional form of the other bank's demand for deposits*:

$$D_I^i(\rho_I^i) = d_I^i = D_B^{i,\neg i}(P^i(\rho_I^i)). \tag{12}$$

Hence, an equilibrium is the 10-tuple $(k^i, d_I^i, \rho_I^{\neg i}, \rho_B^{i,i}, \rho_B^{i,\neg i})_{i=(H,F)}$ that solves equations 9-11, 7, and 12 for $i = H, F$.

5 Graphical Analysis of Equilibrium

TBD.

6 Policy Experiments

TBD.

References

Yesol Huh and Sebastian Infante. Bond market intermediation and the role of repo. *Journal of Banking & Finance*, 122:105999, 2021.