BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2007-28

Modelling Payments Systems: A Review of the Literature

by Jonathan Chiu and Alexandra Lai

www.bankofcanada.ca

Bank of Canada Working Paper 2007-28

April 2007

Modelling Payments Systems: A Review of the Literature

by

Jonathan Chiu and Alexandra Lai

Monetary and Financial Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
jchiu@bankofcanada.ca
laia@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

The authors thank Neville Arjani, Gregory Caldwell, James Chapman, Alejandro Garcia, Walter Engert, Sharon Kozicki, Césaire Meh, Dinah Maclean, Miguel Molico, and seminar participants for comments and suggestions. We also thank David Chen for research assistance.

Abstract

Payments systems play a fundamental role in an economy by providing the mechanisms through which payments arising from transactions can be settled. The existing literature on the economics of payments systems is large but loosely organized, in that each model uses a distinct set-up and sometimes a distinct equilibrium concept. As a result, it is not easy to generalize how model features are related to model implications. The authors conduct a non-technical survey of the literature and discuss some of these connections. They organize the literature according to three general classes of modelling approaches, and compare those approaches in terms of their strengths and weaknesses. They also describe the policy implications across the three model classes and relate them to the model environment/assumptions. The authors summarize what can be learned from the literature with respect to policy issues and identify areas for future research.

JEL classification: E42, E58, G21
Bank classification: Payment, clearing, and settlement systems

Résumé

Les systèmes de paiement jouent un rôle fondamental dans l'économie en fournissant les mécanismes nécessaires au règlement des transactions. Quoiqu'abondante, la littérature consacrée à l'économie de ces systèmes manque de cohésion sur le plan des modèles, chacun d'eux reposant sur une configuration distincte, parfois même sur un concept d'équilibre différent. Résultat, l'articulation des liens entre les caractéristiques et les implications des modèles n'est pas facile à généraliser. Les auteurs passent en revue les recherches sur le sujet sans s'attarder aux aspects techniques et analysent certains des rapports mis en évidence. Ils rangent les travaux dans trois grandes catégories selon l'approche suivie en matière de modélisation, puis comparent les forces et les faiblesses des modèles de chaque catégorie. Ils décrivent aussi les implications des divers travaux pour la conception et la réglementation des systèmes de paiement, en faisant le lien avec l'environnement et les hypothèses des modèles. Enfin, les auteurs résument les conclusions tirées de leur examen de la littérature et suggèrent des pistes pour de nouvelles recherches.

Classification JEL : E42, E58, G21
Classification de la Banque : Systèmes de paiement, de compensation et de règlement

1. Introduction

Payments systems play a fundamental role in an economy by providing the mechanisms through which payments arising from transactions can be settled. In a broad sense, payments systems can be defined as "systems of exchange financed by private and/or public liabilities and the institutions that facilitate the clearing and settlement of these instruments" (Lacker 2005). This broad definition encompasses all non-barter exchange in the economy. In a narrow sense, payments systems can refer to the interbank settlement system, which is "a contractual and operational arrangement that banks and other financial institutions use to transfer funds to each other" (Zhou 2000).

In an interbank settlement system, member banks hold funds at a common agent, the settlement institution or clearing house, and settle the payments among themselves by exchanging liabilities of this settlement institution, the settlement asset (CPSS 2003). In practice, large-value payments systems (LVPS) in many countries are operated by central banks, and most payments systems settle in central bank funds. There are three main types of interbank payments systems: net settlement systems, gross settlement systems, and correspondent banking. In deferred net settlement (DNS) systems, payments between members are collected over a given time period and gross payments are netted against each other at settlement time, with only the net balances having to be settled with finality.[1] In real-time gross settlement (RTGS) systems, transactions are cleared and settled on a real-time and gross basis. In correspondent banking arrangements, the correspondent bank acts as a clearing agent by providing payments services for groups of usually smaller or foreign banks (indirect clearers), which do not have cost-effective access to the primary domestic payments systems.

Why should policy-makers care about payments systems? Owing to the large sizes of existing payments systems and the involvement of the central bank in these systems, policy-makers may care about how reliable these systems are, and how central banks may protect themselves from bearing excessive risk.[2] Moreover, the settlement system is policy relevant because it is the platform through which monetary policy is implemented. While policy-makers care about payments systems, guidance provided by economic theory has, until recently, been limited. In particular, standard Walrasian models abstract from the mechanism through which payments take place and thus are not suitable tools for payments system

[1] A payment is final if funds received by a bank are irrevocable.

[2] To give an example, in Canada, average daily transfers in the Large Value Transfer System are 10 per cent of the annual GDP.

modelling. Recently, a large body of work has been developed on the modelling of payments systems. However, most of this work consists of loosely related papers, in that each paper may have a different model set-up and a distinct equilibrium concept. Therefore, it is not easy to generalize how model features are related to model implications. The goal of this literature survey is to provide a non-technical review of the literature and to discuss some of these connections.

What are the key questions the existing literature tries to address? First, researchers ask why payments systems are essential: what are the fundamental frictions that underline the use of payments and settlement arrangements? Second, given that payments systems are essential, who should provide these systems, and what is the government's role in mitigating the fundamental frictions? Third, how should the payments system be designed optimally (concerning, for example, settlement modes, settlement rules, pricing, access, intraday credit, and risk control)?

In section 2, we organize the literature into three general classes of modelling approaches, depending on whether the following three economic relationships are explicitly modelled: bank-client, trading, and interbank settlement. The bank-client relationship refers to the interactions between households and banks to make loans and deposits. The trading relationship refers to the interactions between households to trade goods or financial assets. The settlement relationship refers to the interactions between banks to settle payments generated by the underlying trading among their client households. Type 1 models focus on the interbank settlement and consist of reduced-form, partial-equilibrium analysis; Type 2 models feature the settlement and trading relationships; and Type 3 models feature all three relationships. Both Type 2 and 3 models involve general-equilibrium analysis. We describe and compare these approaches in terms of their strengths and weaknesses, and the trade-off they represent between tractability, realism, and micro foundations.

In section 3, we draw implications from the literature with respect to the fundamental frictions (for example, trading, information, and commitment) that underlie the need for payments and settlement. We also summarize the policy implications across all three classes of models with respect to system provision, pricing, access, tiering, and system design, and relate them to the model's environment and assumptions. Section 4 provides some suggestions on modelling Canadian payments systems and briefly reviews current work at the Bank of Canada in this area. Section 5 summarizes what can be learned from the literature with respect to policy issues and discusses outstanding questions and issues for future research.

2. Modelling Approaches

There is a large body of work on payments systems that we will categorize according to a general framework in this section. We can group these studies generally into the practitioner-oriented literature, in which payments system "simulators" have featured prominently, and economic models of payments systems. The latter will form the basis of this literature review.

Simulator-based studies

An important body of literature on payments system design carried out in various central banks is based on a payments system simulator developed by the Bank of Finland (called BoF-PSS2), and similar simulators developed in-house by other central banks. The simulator is a computer program that incorporates key features of the payments system, such as the settlement mode (net or gross), credit limits, and queuing procedures. It takes as inputs historical (high-frequency) payments submission data by system participants and calculates settlement statistics such as credit positions, liquidity requirements, and settlement delays on a daily or intraday frequency. The simulator was originally developed by the Bank of Finland, and other central banks have contributed to further developments to accurately represent key features of specific national LVPS. For example, the Bank of Canada, the Bank of England, and the Bank of Finland sponsored the incorporation of bilateral credit limits.

Payments system simulators have been an important and useful tool for helping central banks understand the dynamics of payments system operation through the use of settlement statistics, and for conjecturing behavioural responses to rule changes and shocks to the system. The simulators have been used to estimate the effects of marginal changes to system rules and the impact of operational or default events on liquidity usage, credit risks, and settlement efficiency under the assumption that no behavioural changes take place on the part of system participants. Such studies have the potential to improve understanding on how to economize on liquidity, on effective gridlock-resolution mechanisms, and for examining worst-case scenarios to assess systemic and gridlock risks. However, the usefulness of simulators as a policy tool, particularly for major policy changes, is limited by the fact that the behaviour of system participants is not modelled. Having appropriate models of the payments system can enhance the simulator's usefulness.

Economic modelling of payments systems

There is a growing literature that builds on existing economic theories to model the payments system. Unlike the simulator-based studies, this literature aims to endogenize the participants' behaviour and to derive equilibrium effects by allowing participants to optimally respond to exogenous changes. This literature draws on techniques and insights from monetary, banking, and industrial organization theories to develop equilibrium models of the payments system in a coherent framework. This survey will focus on this line of research. An economic model can either be micro founded or reduced form. A micro-founded model explicitly models the underlying trading and banking activities and specifies the fundamental frictions that generate the need for a payments system. A reduced-form model abstracts from the underlying economic activities and focuses on the settlement sector relationship.

Models with micro foundations

One of the emerging fields in the recent literature on payment economics aims to develop a micro-founded model of payments systems. A micro-founded model specifies explicitly the underlying technology, preferences, information structure, communication protocols, and other relevant trading frictions. These explicit assumptions about fundamentals will determine the feasibility of alternative institutions. In such an environment, the roles of different payments institutions in facilitating trades can be analyzed. Given an equilibrium concept and a welfare criterion, one can determine outcomes of alternative institutions and rank the outcomes accordingly.

Having micro foundations is desirable for several reasons. First, a model with micro foundations accounts for interactions between the trading, banking, and settlement sectors. Unlike in a reduced-form model, underlying economic activities are not assumed to be invariant to payments policy interventions. This is particularly desirable for studying major policy changes. An explicit model of the economic role of payments instruments used in exchanges allows us to understand the equilibrium effect of policies on prices and allocations. Second, by specifying the underlying agents' preferences over allocation explicitly and by modelling their decisions explicitly, a micro-founded model provides a natural welfare measure to evaluate alternative policy outcomes, and allows meaningful analysis of the allocation of costs and benefits as a result of policy changes. A reduced-form model, on the other hand, assigns ad hoc weights to those costs and benefits. Third, by highlighting the relevant microeconomic frictions, a micro-founded model can provide us with an "essentiality test," which tells us whether public intervention is essential for achieving desirable outcomes that cannot be achieved by any other private payments mechanism. If public intervention is needed, an explicit model may also allow us to evaluate exactly what sorts of technological, informational, commitment, or enforcement advantages are possessed by a central bank or regulator that allow it to improve the equilibrium outcome.

Zhou (2000) identifies four criteria for a good payments system model. First, it should model the underlying transaction of goods or financial assets for which payment has to be made, so that system design affects the allocation of real resources. Second, the model should treat consumption/investment debt generated by the underlying transaction as distinct from payments debt, which is created only for payment needs. The third and the fourth criteria are that both liquidity needs in settlement and credit risk should be incorporated into the model, preferably endogenously by agents' choices. A micro-founded model is needed to satisfy these criteria.

Micro foundations, of course, come to us at a cost: micro-founded models tend to be more elaborate than reduced-form models. It is often not easy to capture all the institutional details in a micro-founded model and, at the same time, keep the model analytically tractable. A reduced-form model focuses only on the settlement sector, and is often more tractable and allows one to incorporate more institutional details.

In this survey, we will examine both micro-founded models and reduced-form models. Our survey will note some connections between these models, contrast the main strengths and weaknesses of different approaches, relate model features to model implications, and suggest which approaches are more useful for which types of issues.

2.1 A general framework

We first describe a stylized framework useful for classifying existing models of payments systems into three distinct families of models. In this general framework, there is an economy with a set of households, banks, and a central bank. Figure 1 gives an example of how payments are settled in this economy. Household H_1 deposits money ($, issued by the central bank) in bank 1 and holds a bank liability (IOU) to purchase good Y from household H_2. Household H_2 then deposits this IOU in bank 2, which will settle the IOU with bank 1 in outside money. Bank 1 may need to borrow liquidity from the central bank (CB) to settle the payment; IOU_{IC} in the graph denotes an IOU issued by Bank 1 to exchange for central bank liquidity.[3]

The above set-up is designed to capture the three basic relationships in a payments system:

[3]This set-up abstracts from the functional feature that, in practice, payments settlement is often carried out on the books of the central bank.

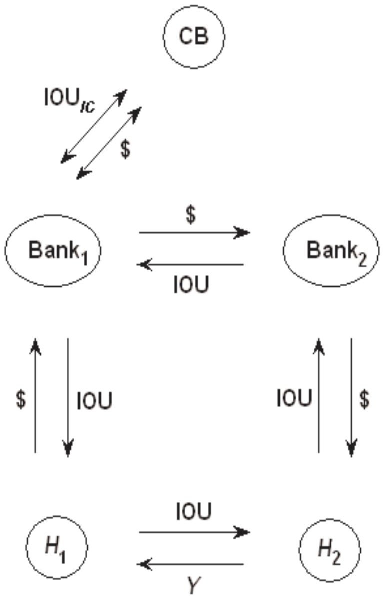

Figure 1: A General Framework

the underlying trading relationship, the bank-client relationship, and the interbank settlement relationship. Conceptually, we can decompose this payments network into three "sectors." Figure 2 graphically summarizes this set-up. In the trading sector, households interact among themselves to trade goods or financial assets. In the banking sector, households contract with banks to make loans and deposits. In the settlement sector, banks interact among themselves for interbank payments settlement, which may involve liquidity provision by the central bank.

This framework allows us to organize the existing literature by identifying the basic structure of a payments system model. Models surveyed in this paper all feature the settlement relationship and some also incorporate the other two relationships. We thus classify models according to the relationships that are explicitly incorporated. We come up with three distinct families of models. The first type features only the settlement relationship, taking the other two relationships as exogenous. The second type incorporates both the settlement and the trading relationships, and the third type fully models all three relationships.

In this section, we will describe these classes of models without going into detail regarding the questions these models have been used to address. Models can also differ along other dimensions (for example, by equilibrium concept or by specific features such as information structure), and we identify these as extensions or variations within our classification scheme. We discuss the models' policy implications in section 3.

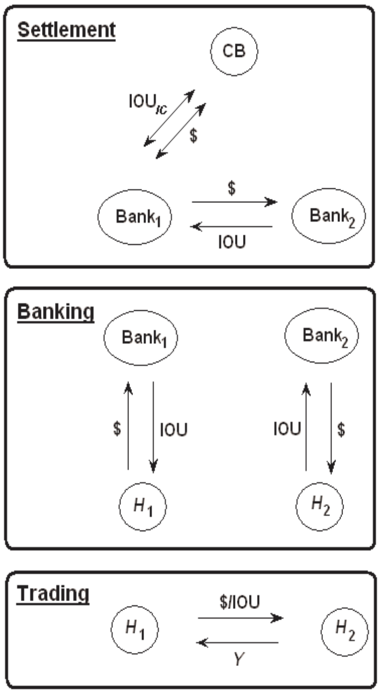

Figure 2: Interbank Settlement, Bank-Client, and Trading Relationships

2.2 Type 1 model: interbank settlement relationship

The first class of models focuses only on the interbank settlement sector, and abstracts from the underlying bank-client and trading relationships. Some examples are Furfine and Stehm (1998), Kahn and Roberds (1998), Angelini (1998), and Holthausen and Rochet (2003, 2005). These models are developed to analyze the liquidity-management problem of an individual bank, or the interaction between multiple banks participating in the settlement sector. Typically, a bank in this setting is given a reduced-form objective function and has to decide on, for example, its portfolio composition, demand for central bank credit, the timing and the size of payments, and whether to default. An arbitrary preference for the central bank can then be used to derive the optimal policy.

2.2.1 A simple example

To highlight the features of this model class, we present a simplified version of Furfine and Stehm (1998) to capture the static liquidity-management decision of an individual bank. Consider a bank that settles its payments in an RTGS system. It is faced with an exogenous and uncertain amount of payments demand, $\tilde{T}$. Uncertainty arises because some payment flows are driven by the bank's customers and therefore cannot be perfectly predicted. Define Z to be the bank's desired use of intraday, central-bank-provided credit. If the intraday credit is costless and unlimited, a bank may choose to process each and every payment as it arrives, borrowing intraday credit from the central bank whenever the payment amount exceeds its reserve account balance. Define $B(\tilde{T})$ to be the amount of credit that the bank would use when intraday credit is both unlimited and free.

When intraday credit is costly, a bank will choose to economize on its liquidity by using less intraday credit than $B(\tilde{T})$ to process a payment volume, $\tilde{T}$. Furfine and Stehm assume that liquidity-management costs are incurred when the bank tries to reduce intraday credit usage below $B(\tilde{T})$ (e.g., payment delay). This is captured by a liquidity-management cost function $L(B(\tilde{T}) - Z)$ that is an increasing and convex function of its argument.

The central bank in the model is assumed to have three tools at its disposal. First, there is a price, p, on all intraday borrowing. Second, a fixed cost, q, is imposed if the bank uses more intraday credit than a quantity limit, $\bar{Z}$.[4] Third, a share, h, of the intraday credit has to be backed by collateral. The collateral cost is captured by a function $C(Zh)$. The bank's objective function is to solve

$$\min_{Z} E[L(B(\tilde{T}) - Z) + pZ + C(Zh) + qI_{Z>\bar{Z}}],$$

where $I_{\bar{Z}>Z}$ is an indicator function such that it equals one if $\bar{Z} > Z$, and equals 0 otherwise. A bank chooses the amount of borrowing to balance the trade-off between liquidity-management cost and intraday credit cost.

After solving the bank's problem, Furfine and Stehm then investigate the central bank's problem by assigning an ad hoc objective function to the central bank. A central bank needs to choose its intraday credit provision policy (p, $\bar{Z}$, q, and h) to minimize the expected value of the private cost of commercial banks and the social costs of gridlock, collateralized credit extension, uncollateralized credit extension, and policy implementation.

In conclusion, this model abstracts from explicit trading and bank-client relationships, and examines in detail an individual bank's reserve management decision in response to changes in settlement policies.

[4]When q is infinite, banks are not allowed to borrow more than $\bar{Z}$.

2.2.2 Extensions and variants

The basic set-up introduced above highlights the reserve management problem of an individual bank. There is a line of literature that generalizes this basic set-up in several dimensions: endogenous default, strategic interactions between banks, and multiple payments systems.

An individual bank

The simple example discussed above is a purely static liquidity-management problem. Kahn and Roberds (1998) and Lester (2005) examine the dynamic liquidity-management problem of a bank facing stochastic payment flows throughout the day. One can also explicitly introduce the possibility of default. The partial-equilibrium model developed in Lester (2005) introduces an exogenous probability of default, whereas Kahn and Roberds (1998) allow the default decision to be endogenous. It is assumed that a bank's default will generate a social cost, which is an increasing function of outstanding liabilities of the defaulting bank. In this setting, banks do not fully internalize the social benefit of keeping reserves, and the central bank faces a trade-off between the liquidity cost and the default risk. On the one hand, a tighter intraday credit policy encourages banks to keep more reserves and thus raises the liquidity cost of the banks. On the other hand, more conservative reserve management induced by a tight credit policy can reduce the size of potential default. In order to determine the desirable policy, arbitrary weights have to be assigned to these two objectives (or to other cost/risk measures that policy-makers care about).

Multiple banks

Other models study the strategic interaction among banks in a payments system and hence explicitly model the behaviour of multiple banks with respect to their investment in reserves (Chakravorti 2000), or with respect to the timing of their payments to each other (Angelini 1998; Bech and Garratt 2003). Chakravorti (2000) constructs a model of multilateral net settlement with multiple banks to study the trade-off between systemic risk and the cost of interbank transfers. The model assumes exogenous default and is used to derive a condition that determines the maximum number of bank defaults that can be sustained without a stoppage in the clearance and settlement process. Angelini (1998) and Bech and Garratt (2006) develop an RTGS model with two banks and endogenize banks' choice of the timing of payment. The model studies the effect of costly daylight liquidity on banks' incentive to postpone outgoing payments and/or to keep excess precautionary reserves.

Multiple systems

Holthausen and Ronde (2002), and Holthausen and Rochet (2003, 2005), extend this line of inquiry to questions regarding access policy and optimal pricing of payment services by payments system operators by considering competition between two payments systems. In Holthausen and Ronde (2002), a gross and a net settlement system coexist and banks in two countries transfer payments to each other through one of those systems. National regulators determine national access policies regarding the use of the net settlement system, while the use of the gross settlement system is unrestricted. Pricing by system operators is not examined here, but the two papers by Holthausen and Rochet (2003, 2005) consider the optimal pricing of payment services by a publicly operated payments system. The market structures studied include: (i) a public monopoly, (ii) a mixed duopoly (one public and one private system), and (iii) two national public systems in a two-country model. These models abstract from liquidity decisions by banks (and ignore the liquidity costs to banks of using alternative systems), but focus on banks' choices of payments system to send payments through.

2.2.3 Discussion

A Type 1 model is able to endogenize banks' behaviour, the interaction between banks, and policy effects in a partial-equilibrium setting. By construction, the model takes as exogenous the use of payment instruments, the need for settlement, payment flows, and the bank-client relationship. The main questions that can be addressed concern the cost-risk trade-off and the policy effects on behaviour. However, this type of model is not suitable to address questions concerning why certain payments and settlement arrangement is essential to an economy, how a policy affects trading and banking relationships, or concerning the effects of monetary policy, welfare analysis, and optimal system design.

Strengths

These partial-equilibrium models can provide a very useful tool to capture the main trade-offs faced by banks and how payments system policy can affect these trade-offs. Moreover, these models are able to capture many institutional details and at the same time remain analytically tractable.

Weaknesses

Type 1 models have a number of weaknesses. First, they are unable to analyze the full policy implications for the allocation of resources and prices in the economy. Failing to satisfy Zhou's (2000) first criterion, these models are inadequate if we want to study the full equilibrium effect of payments system policy on the economy. Without actually modelling the underlying banking and trading activities, a partial-equilibrium model at best can only predict the policy implications on variables determined within the system; it cannot capture the ultimate effects on the underlying activities that generate the need for settlement.

Second, policy implications may not be robust, since key parameters in a reduced-form model capturing the underlying economic activities need not be invariant to policy intervention. This class of models makes shortcuts by imposing ad hoc functional forms or parameter values (for example, on the bank's return function and the process of underlying payment flows) for analytical convenience. In reality, these reduced forms and values are likely to change as agents in the trading and banking sectors modify their behaviour to respond to changes in policy, information technology, and other aspects of the environment. As a result, this class of models may not provide an ideal framework for studying effects of major policy changes, optimal institutional design, or long-run system development.

Third, the welfare analysis is inconclusive. Since the underlying economic agents are absent in the model, any attempt at welfare analysis within the partial-equilibrium setting would require arbitrary weights assigned to various measures (such as costs and risks).

Finally, this type of model is not a desirable framework for studying optimal payments system design. Without explicitly specifying what microeconomic frictions are present and endogenizing the essentiality of settlement and various payment instruments, it is hard to tell what payments institutions are technologically and incentive feasible.

2.3 Type 2 model: interbank settlement and trading relationships

Another class of models abstracts from the bank-client relationship and builds a general-equilibrium model consisting of trading and settlement sectors. Examples of this class of models include Freeman (1996b), Kahn and Roberds (2001), and Koeppl, Monnet, and Temzelides (2006).

A workhorse model is developed by Freeman (1996a, 1996b).[5] The idea is to use an

[5]The intuition of the optimality of elastic money supply in Freeman (1996 a,b) stems from a rich literature on free banking and the real-bills doctrine. See Sargent and Wallace (1982) and Champ, Smith, and

overlapping-generations model, so that outside money has value in equilibrium and its use is essential for efficiency, and to posit some heterogeneity among agents within each generational cohort in order to provide an incentive and efficiency rationale for different types of transactions, such as exchanges of goods for IOUs and the settlement of those IOUs.

The basic Freeman model consists of only a commodity trading period and a settlement period. In the commodity trading sector, some agents need to purchase consumption goods before selling production goods (debtors), while others sell goods before purchasing goods (creditors). As a result, debtors need to pay by credit (issuing IOUs), which has to be repaid in the settlement period. In the interbank market, there is a shortage of liquidity, owing to the stochastic nature of settlement timing. In particular, it is assumed that some creditors have to leave the interbank market before they can be fully repaid. As a result, creditors can gain from trading IOUs among themselves. Freeman shows that there is a potential welfare-improving role for central bank intraday credit. In this setting, the effects of policy intervention on consumption, production, and welfare are explicit. A brief outline of the set-up of the Freeman model is given below. Readers not interested in technical details can skip the following subsection without any loss of continuity.

2.3.1 A simple example

The original Freeman model focuses only on the trading and settlement sectors; there is no explicit role for banks. In order to introduce intermediaries, this section reinterprets the Freeman model in such a way that each household owns a "bank" to manage its payment settlement.

We consider a discrete-time, infinite-horizon model with periods denoted by $t = 0, 1, 2....$ Each period is subdivided into two subperiods: day ($s = 1$) and night ($s = 2$). Therefore, the timing is denoted as (s, t), with $s = 1, 2$ and $t = 0, 1, 2, ...$ As shown in Figure 3, the trading sector operates in subperiods $s = 1$ and $s = 2$, while the settlement sector operates in subperiod $s = 1$ only. We use an overlapping-generations model to endogenize the value of money: each household lives for two periods and uses money as a store of value. To endogenize the need for borrowing and lending, we assume that there are two types of households: creditors and debtors. Debtors want early consumption and thus need to borrow, while creditors want late consumption and thus are willing to lend. Specifically, generation t debtors have endowments in time $(2, t)$ but want to consume in time $(1, t)$, while generation t creditors have endowments in time $(1, t)$ but want to consume in time $(2, t + 1)$.

Williamson (1996) for references.

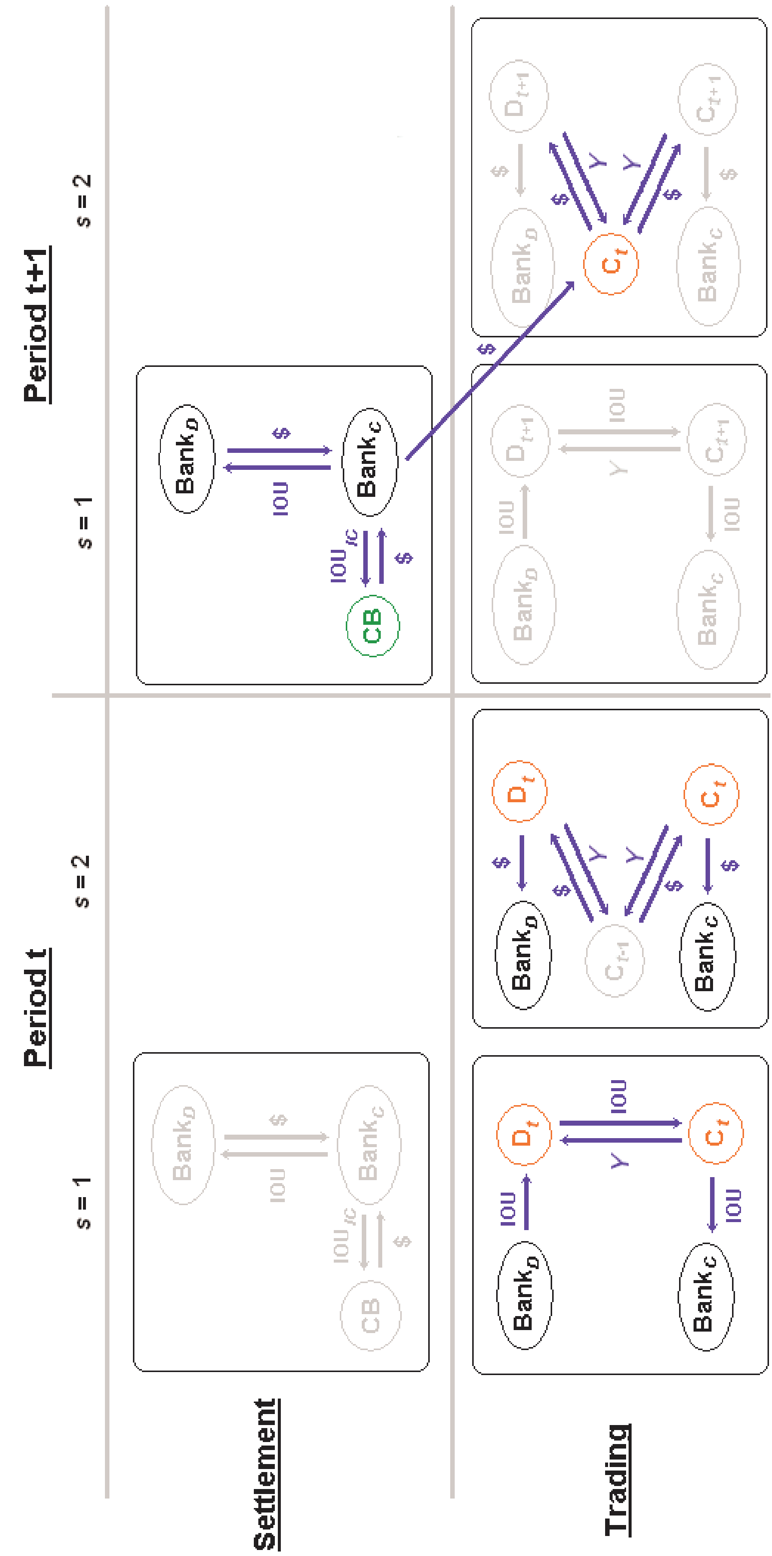

Figure 3: A Type 2 Model

Freeman assumes an "island" economy set-up with spatial separation to generate some specific trading patterns. We first consider debtors in the time $(1, t)$ trading sector (i.e., the first graph in the bottom row of Figure 3). A young debtor born in time-$(1, t)$, denoted by D_t in the graph, wants to consume in $(1, t)$, but without endowment. Therefore, the debtor has to buy consumption from young creditors in $(1, t)$, denoted by C_t in the graph, with an IOU issued by the debtor's own bank. We then consider debtors in the time $(2, t)$ trading sector (i.e., the second graph in the bottom row of Figure 3). In $(2, t)$, a young debtor sells her endowment to old creditors, denoted by C_{t-1}, for money and pays it to her bank. Next, we move to the time $(1, t+1)$ settlement sector (i.e., the second graph in the top row). Here, the debtor's bank uses this money earned by the debtor to redeem the IOU held by the creditor's bank in the settlement sector in time $(1, t+1)$.

We next consider creditors in the time $(1, t)$ and $(2, t)$ trading sectors (i.e., the first two graphs in the bottom row of Figure 3). A young creditor born in time-$(1, t)$ has an endowment of one unit of consumption good, which can be sold in $(1, t)$ and $(2, t)$. This creditor wants to consume when old in $(2, t+1)$. Therefore, a creditor will sell part of his endowment (to young debtors) in $(1, t)$ for IOUs, and then sell the remaining endowments (to old creditors) in $(2, t)$ for money. The creditor will then deposit the IOUs and the money in the bank. In the time $(1, t+1)$ settlement sector (i.e., the second graph in the top row), the bank represents the creditor in receiving repayment from the debtor's bank, and transfers all money balances collected to the creditor for consumption. Finally, in $(2, t+1)$, the creditor uses the money balances to purchase goods from generation $t+1$ for consumption (i.e., the last graph in the bottom row).[6]

Denote the price of the consumption good in period t by P_t. A young creditor has to decide what fraction, α, of his endowment will be sold to debtors for an IOU (with a nominal value denoted by L); the remaining fraction, $1-\alpha$, will be sold to old creditors for a money balance (denoted by Z). A debtor promises to repay her creditor's bank (through the creditor's own bank) in the settlement subperiod, and each IOU carries a nominal interest rate, r. The interest income will be earned by the creditors, the owners of their banks.

We can introduce liquidity risk in the settlement sector by assuming that the timings of "due-to's" and "due-from's" of a creditor's bank are not synchronized.[7] With a probability p, a creditor's bank has to transfer to the creditor before receiving repayment from the debtor's

[6] This set-up follows Ghwee (2006) to modify the Freeman model to incorporate reserve holding.

[7] One may also introduce default risk by assuming that a debtor's bank will not redeem the IOU with some probability.

bank, and thus faces a liquidity shortage. In this case, the bank has to borrow intraday credit from the central bank at a rate i, with the intraday loan being backed up by the forthcoming loan repayment, $L(1+r)$. In this case, the creditor will receive a transfer $T_1 = Z + \frac{L(1+r)}{1+i}$. With a probability $1-p$, there is no liquidity shortage and the bank can transfer to its creditor after receiving repayment from its debtor's bank. In this case, the creditor will receive a transfer $T_2 = Z + L(1+r)$.

The net worth of the bank at the end of the settlement round is

$$NW = \begin{cases} NW_1 = L(1+r) - (T_1 - Z) - \max\{T_1 - Z, 0\}i & \text{with probability } p, \\ NW_2 = L(1+r) - (T_2 - Z) & \text{with probability } 1-p. \end{cases}$$

Here, a creditor's bank is just modelled as a representative of the creditor in the settlement market. Since a creditor cares only about his consumption in the subperiod $(2, t+1)$, the net worth of the bank will transfer to the creditors for consumption use. Let the utility function of a creditor be $U(C)$. Denote the $(2, t+1)$ consumption of a creditor as $C^1_{2,t+1}$ if there is a liquidity shortage, and as $C^2_{2,t+1}$ if there is no liquidity shortage. Given (P_t, P_{t+1}, r_t, i_t), a creditor chooses α to maximize:

$$pU(C^1_{2,t+1}) + (1-p)U(C^2_{2,t+1}),$$

subject to

$$\begin{aligned} C^1_{2,t+1} &= \frac{1}{P_{t+1}}[(1-\alpha)P_t + \frac{\alpha P_t(1+r_t)}{1+i_t}], \\ C^2_{2,t+1} &= \frac{1}{P_{t+1}}\left[(1-\alpha)P_t + \alpha P_t(1+r_t)\right]. \end{aligned}$$

Therefore, the creditors' utility-maximization problem takes into account the banks' liquidity-management problem. A positive intraday interest rate, i, will imply that creditors without liquidity shortages consume more than those with liquidity shortages. If creditors are risk-averse, consumption variation among ex ante identical creditors may be suboptimal.

Given i, we can solve for the equilibrium prices and interest rates as well as the consumption profiles for debtors and creditors.

In conclusion, by modelling the underlying trading, the Freeman model explicitly specifies the microeconomic frictions (e.g., timing frictions, spatial separation, imperfect information regarding trading histories) that generate the need for this specific payments and settlement system: the coexistence of monetary and credit trade and the use of money for settlement. This simple model can capture how changes to payments system policy will affect the allocations and prices in the underlying trading sector, which in turn will feed back to the system by affecting the incentives of banks and the payment flows.

2.3.2 Extensions and variants

The literature based on Freeman (1996b) is primarily built to analyze the welfare-improving role of intraday credit policy in various settings. Some papers use the mechanism design approach to derive the optimal payments institution. Other models, incorporating both trading and settlement sectors, are developed to compare the relative merits of different settlement rules.

Endogenous default

Freeman (1999) modifies the basic model above to incorporate default risk, but the timing of agents' arrivals and departures, as well as whether debtors default, are still exogenous, rather than endogenously chosen by agents. Therefore, the model is not suited for studying the potential moral-hazard problem induced by free or low-cost provision of intraday settlement liquidity, and the possible policies to offset it. Without modelling the decision of whether to default, we cannot discuss the effect of settlement rules on incentives to default, and how default decisions may be strategically interdependent and may thus threaten the stability of the payments system. Some efforts have been made to endogenize agents' default decisions and portfolio quality choices (Martin 2004; Mills 2006). Martin considers the moral-hazard problem by allowing endogenous portfolio choice. He studies the impact of different intraday credit policies, including interest rates, collateral requirements, and net debit caps. Mills allows agents to have the choice to strategically default, and assumes that a technology of costly enforcement is available to the central bank.

Mechanism design

The Freeman model has micro foundations in that the model tries to endogenize the value of the payments system by explicitly specifying the frictions that impede the functioning of markets. This allows us to study the optimal institution in a mechanism design approach, which has been widely used in modern monetary theory (for example, Kocherlakota 1998). One advantage of the mechanism design approach is that it takes a stand on which policies are feasible. The idea is that the features that make a payments institution essential will have implications for the set of policies that are available to the central bank. It is also a valuable tool for understanding what features of an environment expose the payments system to certain risks, and what institutional policies can best address those risks. Recent papers using this approach include Martin (2004), Mills (2004) and Koeppl, Monnet, and Temzelides (2006). Mills (2004) examines a Freeman model; he shows that the settlement stage works as an exogenous (perfect) enforcement device, and that therefore outside money is not needed for settlement, and that central bank provision of intraday credit is not needed for solving liquidity problems. Koeppl, Monnet, and Temzelides (2006) develop a model based on Lagos and Wright (2005), and introduce a centralized market as a settlement sector. They consider an environment with a lack of commitment and private information about production opportunities. They find that the settlement stage is essential: first-best allocation is implementable if and only if settlement is introduced, provided that it takes place with a sufficiently high frequency.

Other extensions of the Freeman model

Much research builds on Freeman (1999) and extends the Freeman model to study different aspects of payments systems. Zhou (2000) introduces money growth into the model. Ghwee (2006) studies the distributional effect of monetary policy. Green (1999) studies the feasibility of private provision of intraday liquidity in the settlement sector. Freeman (1996a) examines the use of private money as a means to settle debt in this environment. Freeman (1999) introduces aggregate uncertainty of default risk to study the risk-sharing role of liquidity provision. Another recent development is to generalize the Freeman set-up to a two-country version of the model to study international settlement (Chapman 2006; Fujiki 2006).

Most of the Freeman-based literature abstracts from the frequency of settlement and focuses instead on intraday credit policy. In the basic Freeman model, agents are either pure debtors or creditors, and thus there is no distinction between net and gross settlement. Freeman and Hernandez-Verme (2004) modify the trading pattern in Freeman's model so that all agents are both debtors and creditors, to evaluate the relative merits of different settlement rules. In a similar fashion, Kahn, McAndrews, and Roberds (2003) build a two-sector general-equilibrium model to study the gridlock problem in net and in gross settlements.

One criticism of the Freeman models arises from their "island" economy set-up. It is not immediately clear what types of frictions are captured by these "island" features (geographical, timing, technological, or institutional), or how robust the implications are when we modify the island set-up. An ideal approach would be to incorporate the Freeman set-up and the standard monetary models into a unified framework, where geographical isolation in an island economy model can be recast by appealing to other features that lend themselves to better economic interpretation. For example, Nosal and Rocheteau (2006) use a random matching model to capture the essential features of Freeman (1996b), and illustrate that the settlement system can be easily incorporated into a standard micro-founded monetary model. Berentsen and Monnet (2005) reinterpret the monetary model of Lagos and Wright (2005) and the banking model of Berentsen, Camera, and Waller (2005) as models of a payments system. As in Freeman's model, both the trading and the settlement sectors are incorporated. But preference/productivity shocks and a centralized market are introduced to replace "islands" and overlapping generations, to endogenize the value of money, and to gain tractability.

2.3.3 Discussion

By incorporating the trading sector, a Type 2 model is able to study the effect of payments and monetary policies on the underlying trading. Unlike a Type 1 model, this class of models is micro founded and can address fundamental questions like why settlement is needed; it can also be used to conduct welfare analyses and examine payments system design. Because of their construction, these models cannot examine the interaction between the settlement and the banking sector as well as the bank-client relationship.

Strengths

This class of models has the following strengths:

First, by specifying the underlying preferences and technologies, the model provides a natural definition of welfare: a weighted average of expected utilities.

Second, with the underlying trading modelled explicitly, the model permits a transparent analysis of the effects of payments system policy on economic activity, welfare, and distribution, as well as the use of different payment instruments.

Third, the microeconomic frictions specified in the model suggest the types of policy that are feasible, making the study of optimal institutional design in this environment meaningful.

Weaknesses

One limitation of the literature on this class of models is that the role of banks is not modelled. Hence, these models are not suitable for answering questions related to how payments system policy may affect the banking structure, and how the banking structure may affect policy implications, since the models do not distinguish between trades for banks' customers and trades for the banks' own account. By modelling banks and households as a single agent, these models ignore the fact that they may have different incentives and attitudes towards risk. Another potential weakness of these models is that the set-up usually is too stylized to capture certain institutional details of a settlement system.

2.4 Type 3 model: interbank settlement, trading, and bank-client relationships

Another class of models attempts to build a full-blown model consisting of the trading, banking, and interbank settlement sectors, to study the general-equilibrium effects of payments and monetary policies. Some examples of this class of models are Freixas and Parigi (1998), Lester (2005), and McAndrews and Roberds (1995). In general, banks are motivated by their issuance of liabilities as payment instruments, and by their provision of settlement services. In some cases, banks also provide risk sharing among risk-averse customers, à la Diamond and Dybvig (1983). Lester (2005) formally introduces money into an infinite-horizon model and thus can study the money-creation role of banks. Freixas and Parigi (1998) and McAndrews and Roberds (1995) study a three-period model and abstract from the use of money.

In the following section, we will discuss a generalized version of the Lester (2005) model. The basic framework is as follows. In the trading sector, owing to anonymity, the lack of commitment, and the lack of double coincidence of wants, goods trading requires the use of certain media of exchange. As a result, in the banking sector, banks accept households' deposits and issue bank liability, which will be circulated as a medium of exchange in the trading sector. In the settlement sector, a bank transfers reserves to other banks for settlement as their depositors trade with other banks' depositors in the trading sector. A brief outline of the set-up of the Lester model is given below. Readers not interested in technical details can skip the following subsection without any loss of continuity.

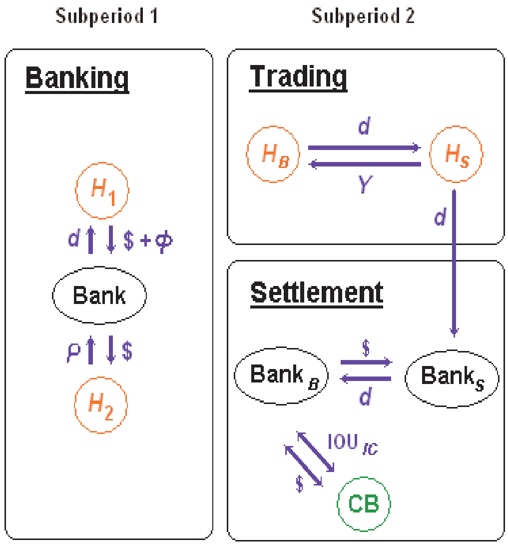

Figure 4: A Type 3 Model

2.4.1 A simple example

As shown in Figure 4, each period is divided into two subperiods. The banking sector opens in the first subperiod in which banks and households interact to make loans and deposits. In the second subperiod, the trading and the settlement sectors open simultaneously. The banking sector in the first subperiod is similar to the framework modelled in He, Huang, and Wright (2005). In this sector, a large number of identical banks serve households by managing their deposit accounts and making loans to them. To gain tractability, we assume the existence of a numeraire good in this sector. The utility of consuming X units and producing Y units of this numeraire good is $U(X) = X$ and $C(Y) = Y$, respectively.

Money is indivisible: households can hold either zero or one unit of money. There is a [0,1] continuum of households, and a fraction, M, are initially endowed with one unit of fiat money issued by the central bank. The left-side graph in Figure 4 describes the banking sector. Here, a money-holding household (H_1 in the graph) can choose to deposit money at the bank in exchange for a "check," d. The bank charges a cost, ϕ (in terms of the numeraire good), for this service. We will focus on equilibrium with $\phi < 0$, so that ϕ works as interest earned on deposits, and all money-holding households choose to deposit money in their banks. An agent without money (H_2 in the graph) can take out a loan for an upfront fee, ρ (in terms of the numeraire good). Also, a bank has to incur a fixed cost, a, in managing a deposit account. Denote the reserve ratio chosen by banks (discussed below) as α. In equilibrium, a fraction, $M_1 = \frac{M}{\alpha}$, will enter the trading sector holding checks as buyers, while the remaining $1 - M_1$ will enter without checks as sellers. As a result, the total money stock (checks) circulating in the trading sector is M_1.

Denote r as the rate-of-time preference. When a bank takes a deposit, it charges ϕ and earns revenue $(1-\alpha)r\rho$ in loans, given the reserve ratio, α, chosen by the bank. Also, let $h(\alpha)$ be the expected cost of intraday borrowing per unit of deposit for a bank with a reserve ratio α.[8] The expected profits of a bank with D units of deposit are given by

$$\pi(\alpha) = D(\phi + (1-\alpha)r\rho - a) - Dh(\alpha).$$

Given the prices, a bank chooses α to maximize its profit. If $h(.)$ is decreasing in α, then a bank faces a trade-off between liquidity costs and loan returns. The free-entry condition in the trading sector implies $\pi = 0$ in equilibrium.

The right-side graphs in Figure 4 describe the trading and settlement sectors. The trading sector is a decentralized market. In this market, households are subject to bilateral random matching, and meetings occur at T discrete points in time throughout the second subperiod. Each household will randomly match with another household, who may be a depositor at the same bank or at a different bank. The probability of a double coincidence of wants is zero. When a buyer meets a seller, there is probability σ that he likes the seller's product. In this case, the seller produces the good at cost c in exchange for the buyer's check, and the buyer immediately consumes the good and receives utility u. After a seller sells her good, she leaves the market and deposits the received check at her bank. The buyer's bank will transfer one unit of reserve to the seller's bank in the settlement sector. When the seller's bank receives payment finalization, the seller's account is credited one unit and she becomes a buyer the following day.

We use V_0 and V_1, respectively, to denote the value functions of a household with zero money holdings and a household with one unit of money:

$$\begin{aligned} rV_0 &= M_1\sigma(V_1 - V_0 - c), \\ rV_1 &= (1-M_1)\sigma(u + V_0 - V_1) - \phi. \end{aligned}$$

[8]See Lester (2005) for the derivation of this cost function.

A seller (without money) will produce at a cost c and sell with probability σM_1, and becomes a buyer in the following period. Similarly, a buyer (with money) will buy and consume (deriving utility u) with probability $\sigma(1-M_1)$, and becomes a seller in the following period. We assume that sellers deposit the cheques to their own banks immediately after the transaction takes place. If the seller and the buyer belong to the same bank, then no interbank transfer is needed. If they belong to different banks, then the buyer's bank needs to make a transfer of reserve to the seller's bank. Since the timing of transactions is random, a bank accrues due-to's and due-from's throughout the second subperiod as its depositors trade with other banks' depositors in the trading sector. A bank's net liabilities at any time are equal to the difference between due-to's and due-from's. As in the previous sections, should a bank's obligations exceed its reserves, it must borrow the difference from an intraday credit facility provided by the central bank. This liquidity cost is captured by the function $h(\alpha)$ mentioned above.

This model endogenizes the provision of settlement services by banks in the banking sector and the circulation of bank liabilities as media of exchange in the trading sector. Lester (2005) introduces the possibility of exogenous default into this set-up and examines the optimal settlement frequency and the optimal intraday credit policy. Lester shows that, since the frequency of settlement and the costly provision of liquidity are both levers that can be employed to reduce the risks associated with default, the optimal settlement system requires joint consideration of both settlement frequency and intraday borrowing practices. The optimal use of these two levers will depend heavily on the underlying parameters of the economy (probability of solvency and liquidity shocks, as well as the costs of recovering from a default). When there is sufficiently high risk, optimal design requires a combination of high-frequency settlement and costly provision of liquidity. When risk is small, optimal design requires less-frequent settlement, and free intraday liquidity may be optimal.

2.4.2 Extensions and variants

Diamond and Dybvig banks

Lester (2005) bases his model on He, Huang, and Wright (2005) to motivate banks as issuers of payment instruments and to study the role that banks play in money creation. Another line of literature, based on Diamond and Dybvig (1983), models the banks' function of providing liquidity insurance. By extending the Diamond and Dybvig model of banks to include a spatial dimension, a role for bank-issued claims as a means of payment arises naturally. As Lacker and Weinberg (2003) point out: "If a role of banks is to provide agents with insurance against preference shocks and if those shocks have a spatial as well as a temporal dimension in an environment with imperfect and costly information, then an important function of banking contracts will be to economize on communication while retaining the incentive compatibility characteristics necessary to implement optimal consumption allocations."

In Freixas and Parigi (1998) and McAndrews and Roberds (1995), the banking sector is built on the Diamond and Dybvig (1983) structure, which insures its depositors against preference shocks dictating whether they are early or late (impatient or patient) consumers through balance sheet intermediation (taking demand deposits and investing in long-term assets). It does this by adding distinct locations (location represents any way that banks are differentiated from each other) with banks that operate in one location only, and allowing consumers to be subject to a location shock. Some depositors travel to a different location and have to consume at that location. If travellers pay for goods at the new location by cheque, or if they transfer their claims on the home bank to their new location, payments have to be made between banks in different locations. Payments across locations can be made by transferring liquid assets (gross settlement), or by transferring claims on banks that are settled when long-term assets mature (net settlement).

The economy has three time periods ($t = 0, 1, 2$) and a number of banks, each operating in a distinct location. Banks have access to a productive long-term asset that yields a per unit return of R after two periods, or of 1 if the asset is prematurely liquidated. Each bank has depositors with \$1 deposited in the bank. These depositors are identical at date 0, but a random fraction, τ, of them turn out to be early consumers and they have to consume at date 1. The rest are late consumers. Payments across banks (locations) are generated due to the mobility of the early consumers (as in McAndrews and Roberds 1995), or the mobility of late consumers (as in Freixas and Parigi 1998).

In the McAndrews and Roberds model, all early consumers travel to a subset $\mathcal{M}$ location, where they will purchase consumption goods by writing cheques on their deposit accounts. These cheques have to be cleared and settled in the same period, and hence require reserves. The creditor banks (those located in the subset $\mathcal{M}$) thus have excess reserves that they can lend to debtor banks on the interbank market. With sufficient reserves in the system, premature liquidation of long-term assets can be avoided.

In the Freixas and Parigi model, early consumers always withdraw and consume at home, but a fraction, λ, of late consumers are strategic travellers who can choose the location where they consume, while the other $1 - \lambda$ of late consumers are compulsive travellers. Since the number of withdrawals in another location is endogenous, the size of interbank payments is also endogenous. Under gross settlement, liquidation of productive assets occurs in order to make payments across locations, while net settlement economizes on this cost.

Monetary transmission

Another line of research is to study the transmission of monetary shocks in a payments network. Williamson (2006) builds a general-equilibrium monetary model with a rich clearing and settlement structure. Clearing houses provide clearing and settlement services, while agents purchase goods with personal debts and these debts are settled through the clearing houses. The structure of the clearing and settlement system determines how quickly different debts are cleared, which in turn affects relative prices and monetary diffusion.

2.4.3 Discussion

Strengths

These models provide micro foundations for banking by endogenizing the role of banks in issuing liabilities as payment instruments, as well as by providing safekeeping, payment services, and liquidity insurance. Explicit modelling of the banking sector is useful for understanding the payments system because, in the real world, liabilities of financial intermediaries serve as an important means of payment in circulation.

By making an explicit distinction between the decisions of banks and households, this class of models provides a rich environment to study the interaction between the settlement system, trading activities, banking market structure, and payment and monetary policies. In particular, we can study the policy effects on loan and deposit rates, financial contracts, the circulation of bank liabilities, and, ultimately, the underlying economic activities.

Weaknesses

A complete model of trading, banking, and settlement comes at a cost. Some models (for example, Lester 2005) are so rich that analytical tractability may not be feasible. Others are too stylized to capture all the relevant details (for example, money is not modelled in Freixas and Parigi 1998).

This class of models lies in the intersection of theories of money, banking, and settlement. There are inherent tensions between these three building blocks. The existence of a settlement system and banking networks requires centralized arrangements, while the existence of monetary trades requires decentralized interactions. A researcher needs to carefully balance these elements to endogenize the value of the whole payments system. More work has to be done to analyze the essentiality of payments, banking, and settlement in such an environment.

2.5 The choice of modelling approach

The main lesson we have learned from the above discussion is that a payments system arises as a result of fundamental frictions (trading, information, commitment). Therefore, policy analysis should take these frictions seriously into account.[9] Moreover, system participants' behaviour should not be taken as invariant to policy changes. In order to fully understand policy effects, we need to better model the underlying banking and trading sectors together with the settlement sector. However, the incorporation of micro foundations can significantly increase the complexity of the model. At the extreme, micro-founded models can require computational solutions. This increase in complexity has to be balanced against a potential loss of other desirable characteristics.

In principle, an ideal model should be analytically tractable, micro founded, and able to capture essential institutional details. Analytical tractability makes the intuition transparent (so that the model is not a black box), which is important for understanding the results arising from the model. Micro foundation allows us to study the interaction between the settlement sector and underlying economic activities, as well as conduct meaningful welfare analysis and payments system design. A model should also provide a good characterization of realistic markets by capturing certain institutional details to make it relevant, and to allow for easy interpretation. In reality, model selection involves a trade-off between tractability, micro foundation, and realism.

The Type 1 models are suitable for analyzing how small policy changes may affect the behaviour of individual banks and interactions between banks. Also, they are a better choice when a model needs to capture fine institutional details of the payments system. Type 2 and Type 3 models are suitable for deriving the equilibrium effect of major policy changes, particularly when we are interested in the responses of the underlying trading and banking activities, as well as conducting welfare analysis and payments system design. The Type 3 models are an obvious choice when our model needs to capture a realistic banking structure.

[9]For example, in a micro-founded model, any policy that requires perfect information and perfect commitment for implementation should be considered infeasible.

3. Model Implications

In this section, we focus on the implications of the following three questions for the literature:

(i) What are the fundamental frictions that underlie the use of payments and settlement arrangements?

(ii) What is the role of central banks or regulators with respect to system provision, pricing, and access?

(iii) What is the optimal design of payments system?

(i) Fundamental frictions

The first fundamental question in the area of payments system research is what frictions in the economy underlie the choice of payments instruments and the structure of payments mechanism? Recent developments in monetary economics provide the primary insight that certain deviations from the standard Walrasian environment——such as infrequent bilateral trading, a lack of double coincidence of wants, a lack of commitment, and private information concerning trading histories——give rise to the use of certain payments instruments such as fiat money. In this environment, money is "essential" in that it improves the efficiency of resource allocation compared with a world without money. Periodic settlement also arises to mitigate these frictions. Therefore, any meaningful analysis of various payments system policies should be performed in a context that is consistent with these fundamental frictions. We elaborate on these frictions in subsection 3.1.

(ii) Potential roles of government

Who should provide the mechanism and instruments for facilitating payments and effecting the settlement of payment obligations? Do the frictions underlying the essentiality of payments and settlement imply a role for public authorities, and in particular for central banks? These payment and settlement services include the provision of the settlement asset, intraday liquidity, and finality of settlement. Additionally, reserve requirements, for purposes of liquidity management, may be enforced by the provider of the payments system. A related issue is how these services should be priced. We examine these issues in subsections 3.2 and 3.3.

(iii) Design of payments system

Finally, how should payments systems, in particular large value payments systems, be designed to mitigate the fundamental frictions, and to trade off the associated efficiency and risk? One may further classify system design questions into the following three topics.

Access and tiering

Should all banks in the economy participate directly in the payments system? In reality, not all banks participate directly in payments systems. Some make use of correspondent services provided by other banks, an arrangement that creates a tiered structure.[10] One can argue that this tiering or pyramiding is a natural outcome of the need for an efficient way to make payments in a world with many banks, since it would be inefficient for every agent in the economy to have an account with every bank. Correspondent banking relationships appear to be very diverse, with these contractual arrangements being system specific and non-standardized.[11] Correspondent banking can concentrate payment exposures among a small group of clearing or correspondent banks. Therefore, the failure of an important clearing agent can affect a large number of banks and have systemic risk implications if correspondent banking relations are costly to form, or if other barriers prevent indirect clearers from easily obtaining another clearing agent. There are also competitive issues that arise with correspondent bank clearing arrangements, since clearing banks simultaneously provide an essential input (settlement services) and compete directly with indirect clearers in the retail market for the provision of payments services. In subsection 3.4, we consider whether access to payments systems should be restricted for some banks, and whether a tiered payments system is an optimal structure.

Settlement rules

Public policy regarding payments system design is concerned with balancing various risks and costs against each other. There are two other categories of issues related to the design of payments systems. The first involves the choice of real-time gross settlement versus deferred net settlement and the frequency of settlement within that mode. In practice and in theory,

[10] At the top of the hierarchy are settlement institutions (for example, central banks), which provide settlement accounts to participants connecting directly to, and clearing directly in, this "first-tier" network. Of these direct clearers, some act as clearing agents that operate a "second-tier" network, providing settlement accounts to downstream institutions that wish to clear and settle payments indirectly in the payments system. We observe tiering in most payments systems, large-value as well as retail, in industrialized countries, although the degree of tiering varies across systems (CPSS 2003).

[11] Each indirect clearer settles bilaterally with its clearing agent, where gross exposures are typically netted. Clearing agents hence provide (usually uncollateralized) daylight overdraft facilities to indirect clearers as part of the contractual arrangement.

these two dimensions of settlement come bundled: gross settlement is typically real time, while net settlement is deferred to specific times during the payment day (for example, the end of the day). Unless otherwise noted, we refer to uncollateralized deferred net settlement and real-time gross settlement when discussing model implications for payments system policy.[12]

Since payment finality is virtually immediate for each transaction that is cleared in RTGS systems, systemic risk from unsettled claims is limited. An RTGS system is, however, relatively costly for members in terms of the investment in liquidity management and the cost of intraday liquidity. Moreover, banks' ability to pay out may depend on the timeliness of incoming payments, but counterparties may have the incentive to delay outgoing payments, which leads to the queuing of payments and can create widespread liquidity problems or, in the extreme case, system gridlock, a situation where no institution is willing to send outgoing payments. Uncollateralized deferred net settlement is generally believed to be more vulnerable to systemic risk compared with gross settlement, since gross exposures that accumulate between settlement times can become very large and default can potentially occur on unsettled claims. Contagion risk exists if the failure of a subset of banks or the failure of one large bank can take down all the banks in the system. However, the netting of reciprocal gross positions between banks, especially under multilateral netting arrangements, can considerably reduce the effective net debit position, and thereby reduce participants' incentives to default and mitigate the undesirable effects of deferred net settlement on systemic risk. Moreover, net settlement involves lower liquidity costs, owing to the relatively infrequent settlements and smaller net positions, and hence can mitigate gridlock risk.

The key trade-off between net and gross settlement is the cost of intraday liquidity and system gridlock associated with gross settlement, and the cost of default and contagion associated with net settlement. Clearly, the trade-off will depend on intraday credit policies and on other system policies, such as risk management and collateral requirements, that affect the cost and size of potential default.

Credit policy and risk controls

The second category of payments system design issues takes as given the settlement rule (net or gross), and determines optimal credit policies and risk controls within that framework. For example, at what cost should intraday liquidity be supplied? Should there be limits on the size of intraday loans or on payments that can accumulate (net debit caps)? Should

[12]We note that important net developed economics have risk controls that mitigate settlement risk. Some even go so far as to guarantee certainty of settlement. One such example is Canada's LVTS.

collateral requirements be used to limit the cost of default? How should losses in the case of a default be shared by the remaining participants in a system? By introducing credit risk to gross settlement and by increasing the cost of participation in net settlement systems while reducing system risk inherent in deferred net settlement, intraday credit policies and risk controls make gross settlement and net settlement systems more similar to each other with respect to the balance of risks and costs that they entail. The implications of the literature for these questions are examined in subsections 3.5 and 3.6.

The choice of a settlement mode (gross versus net), the frequency of settlement (real time versus deferred), loss-sharing rules in the event of a default under net settlement, the provision of intraday liquidity, and risk controls for constraining the amount of credit risk generated in a payments system each serve to mitigate certain risks and costs, but may introduce other risks and costs to the system. Moreover, they can have distributional consequences for members as well as non-members of a payments system. The balance of these risks and costs will have real effects on the economy, and it is important to explicitly model the linkages between the trading, banking, and settlement sectors in order to understand the aggregate welfare effects of payments system policies.

3.1 The need for payments and settlement

What are the fundamental frictions (such as informational or legal imperfections impeding the functioning of markets) that underlie the use of payments and settlement arrangements? The recent literature argues that *limited enforcement* and *limited information* are the two key micro-economic frictions that explain why some observed payment arrangements are essential to an economy. Limited enforcement refers to a situation where some agents can default on their obligations at little or no cost. Limited information refers to a situation where some agents have no or limited knowledge about other agents' current and/or past actions.

To understand what the consequences of these frictions are, it is useful to examine the reason for the circulation of the most commonly used payment instrument – money. Why would a seller be willing to give up valuable goods or services in exchange for an intrinsically worthless piece of paper which does not yield direct consumption or production values? Recent literature points out that the informational and enforcement frictions are indeed necessary for paper money to serve as a payment instrument in the economy.

Consider an economy in which trades are decentralized and all trade meetings are bilateral and infrequent. First, to rule out spot barter exchanges between agents, an absence of double coincidence of wants is assumed: an agent desires the good of his trading partner but his trading partner does not care for the good of the agent at all. This assumption seems extreme, but it works as a simplifying assumption to capture the fact that, for most barter trading, there are asymmetric gains from trade: an agent desires the trading partner's good more than the partner desires the agent's good (consider a meeting between an economist and a cook).

In itself, the lack of double coincidence of wants is not sufficient to make money essential. One needs, in addition, a lack of perfect enforcement and a lack of perfect information. In an ideal world with perfect enforcement and information, all trades could be facilitated by credit arrangements based on trust and reputation, and thus money would have no role to play. In the absence of enforcement and perfect information, however, trust and reputation cannot be maintained, and thus the use of money as a payment instrument can facilitate trade and improve welfare. In particular, by offering money to a seller, a buyer is able to signal that (s)he has supplied goods or services to other agents in the past. At the same time, the seller is willing to accept money because (s)he anticipates that (s)he will be able to use this instrument in the future to communicate the same information to other agents. Money as an information instrument therefore serves to communicate reliably about the buyer's past trading histories. (Townsend 1989; Kocherlakota 1998).

Assumptions about the extent to which knowledge of individual histories is public knowledge play a crucial role in determining the choice of payment instruments. At one extreme, perfect record keeping implies no role for money, as in a standard Arrow-Debreu world. At the other extreme, in a world without any knowledge of individual histories, all trade has to involve outside money and there is no role for credit or inside money. More interesting cases arise when there is some degree of imperfect knowledge of individual histories. Kocherlakota and Wallace (1998) consider an economy with a lag in updating the public record of individual histories. In general, there can be coexistence of outside money and credit. As the updating lag goes to infinite, almost all trade involves the use of outside money and, as the updating lag goes to zero, outside money is not needed. Also, Kocherlakota and Wallace show that shortening the lag with which individual histories are made public knowledge enhances welfare. Cavalcanti and Wallace (1999) consider an environment in which some agents have known histories and others unknown histories. The former work as issuers of inside money, and the latter as users of inside money. One may interpret inside money issuers as banks issuing their liabilities as a medium of exchange. The idea is that issuers of inside money are making promises of some sort—perhaps to redeem inside money—and that agents with

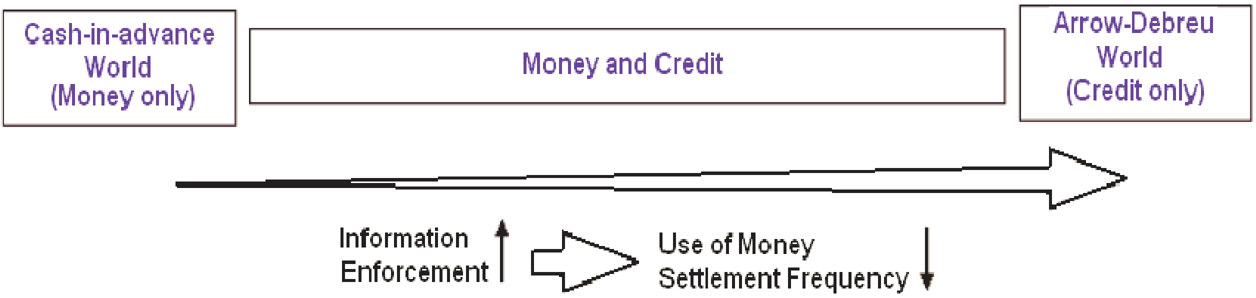

Figure 5: The Need for Payments and Settlement

known histories can be made to keep promises. In this environment, inside money can achieve strictly more allocations than outside money: the use of outside money is more restrictive than is the use of inside money, because current trading opportunities are constrained by past transactions when outside money is used.

Limited information and enforcement are also crucial frictions that make periodic settlement of private liabilities essential.[13] The reason for the need for periodic settlement is not obvious since it merely involves a transfer of settlement assets between participants, without actually improving social welfare by itself. In an ideal world with perfect enforcement and information, default is not a concern, and thus it is efficient to allow agents to accumulate their obligations over time so long as it will be settled at some time in future. In this case, efficient arrangements should not involve periodic settlement other than a lifetime budget constraint. When there are informational and enforcement frictions, however, agents are able to, and may have incentives to, default on obligations. In this environment, periodic settlement helps to reduce the net gain from default by limiting the obligations an agent can accumulate over time. Koeppl, Monnet, and Temzelides (2006) illustrate how periodic settlement with sufficiently high frequency can induce agents to fulfill their payment obligations and improve economic efficiency.

A summary of the above discussion is depicted graphically in Figure 5. Even in an environment with an absence of a double coincidence of wants and infrequent bilateral trading opportunities, money is not needed and credit can be used to facilitate trade if there exists perfect information (record keeping) and perfect enforcement. At this extreme, there is no need for periodic settlement other than a lifetime budget constraint. At the opposite extreme, where no record keeping or enforcement is possible, the use of money is essential and credit arrangement is infeasible. In thsi case, there is immediate settlement on the

[13]For example, credit card transactions have monthly settlement while interbank transactions have daily settlement.

spot of exchange. In between the two extremes, where there is some degree of imperfect information and imperfect enforcement, there is coexistence of money and credit. In general, as information and enforcement improve, we expect the frequency of settlement and the need for money to decline.

3.2 System provision

The need for payments and settlement arises from the existence of fundamental frictions. In this environment, there may be a role for a central bank (and for regulation) with respect to mitigating some of these frictions through commitment and enforcement technologies. This subsection summarizes the characteristics that the literature has identified as required for an agent to be a good provider of various components of a payments system: the provision of the settlement asset, intraday liquidity, and credit insurance through the guarantee of finality, as well as the need to enforce (liquidity) reserve requirements on payments system participants. In most studies, it is clear that it is not necessarily only a public agent who may fulfill these requirements for the welfare-enhancing provision of a payments system. Although a private sector intermediary may be well able to perform the role of issuing a settlement asset and cheaply providing liquidity, a central bank may have several advantages with respect to the commitment not to overissue the asset, to reabsorb the liquidity introduced by intraday needs, and to intervene in times of distress in view of its responsibility for price and financial stability.

Green (2005) and Millard and Saporta (2005) point out other features that can make a central bank the appropriate agent for system provision, including its neutrality with respect to financial institutions, and its creditworthiness. It can also be argued that the provision of settlement accounts to banks is a natural extension of a central bank's traditional role as the government's banker, since commercial banks already hold accounts with the central bank for government transfers. These arguments are further strengthened if there exist economies of scope in the provision of the settlement asset and other public policy objectives, such as the implementation of monetary policy and a concern about systemic risk, that are not internalized by private clearing-house operators. Green and Todd (2001) argue that a central bank's role in the payments system should be limited to providing those services for which it has a comparative advantage—notably, that of providing settlement accounts—and potentially additional services that are justified by economies of scope.

(a) *Provision of outside money as a settlement asset*

Mills (2004) uses a mechanism design approach to study whether outside money is needed for settlement in Freeman's environment. He argues that, if agents have a *technology to issue IOUs*, and there exists an *enforcement authority* to force debtors to redeem their own IOUs, then no outside money is needed for settlement. In this environment, a debtor can first issue personal IOUs to purchase goods, and then he needs to sell his own endowment in exchange for others' IOUs, which will eventually be used to redeem his own in the settlement sector. Mills shows that efficiency can be achieved by this mechanism without the use of outside money. In the absence of this strong enforcement authority, however, outside money will be needed.

(b) *Provision of outside money as intraday liquidity*

Freeman (1996a) considers a clearing house operated by private agents to settle debt and to provide private bank notes to circumvent participants' liquidity shortages. The clearing house's charter is literally a permission to print fiat money and thus enjoy the profits from creating valued money at no cost. An uncontrolled total money stock may result in an inflationary bank note overissue and thus may require regulations.

Green (1999) clarifies that, in Freeman's model, ordinary private agents can assume the role of central bank or clearing house. Private sector intermediaries may issue debt that works like central bank intraday credit. In this case, there is no need for direct participation by a monetary authority. In general, whether efficiency might require a central bank to participate in the payments system depends on the degree to which a central bank and a private intermediary can *promise credibly to reabsorb money* that it issues to facilitate payments, and on whether the commercial law warrants agents' use of debt issued by the private intermediary as a medium of exchange.

(c) *Provision of credit insurance*

Kahn and Roberds (2002) argue that the government can provide credit insurance for participants if they have a different incentive than private agents, and access to cheap liquidity by taxation. In an environment with limited enforcement, private agents need to post costly collateral to commit to settle obligations, particularly during times of distress. However, a government following a "too-big-to-fail" policy can commit itself to intervene during times of distress. The central bank's "promise" can be supported by its ability to tax. The presence of backup liquidity through the central bank affords the members of private payments systems assurances that promised payments will be made, without posting the quantity of collateral that would be necessary to provide such guarantees in a purely private context.

(d) *Enforcing the reserve requirement*

McAndrews and Roberds (1995) develop a model with a role for a payments system operator to enforce the reserve levels of payments system participants in a net settlement system. Assuming that all banks in the economy are members of a clearing house that can observe and enforce reserve levels, the constrained efficient equilibrium entails full insurance across consumers, and zero interbank interest rates. However, the optimal level of reserves under this scheme is not self-enforcing: banks would rather hold lower reserves, since excess reserves earn a zero return when lent out. Given that liquidity levels need to be regulated in such a world, McAndrews and Roberds are silent on whether the clearing house should be a private or public agent. However, one can argue that a central bank has a natural advantage in enforcing reserve levels in a system operated by a central bank that settles in central bank funds.

When there are intermediaries that are not members of the clearing house and, therefore, not subject to the clearing-house-wide reserve requirement, the clearing house is required to provide free liquidity to its member banks in a financial crisis (aggregate liquidity shortage), in order to maintain full insurance across depositors of its member banks in the constrained efficient equilibrium. As before, the role of clearing house can be played by a private or public agent.

3.3 Pricing policies

Holthausen and Rochet (2003) examine the optimal pricing policies of public payments systems, whose costs of operations include a fixed costs and a constant marginal cost, under various market structures: (i) a monopoly public payments system, (ii) a mixed duopoly consisting of a public payments system and a private payments system, and (iii) a duopoly consisting of two national public payments systems. Public payments systems are required, by assumption, satisfy a balanced-budget constraint. The authors find that the that the optimal pricing scheme of the public payments system depends on the market structure, but, in general, takes the form of *two-part pricing with quantity discounts*, so that "more efficient" banks (that submit more payments) pay less than the "less efficient" banks. When the mixed duopoly case is augmented to allow for an unpriced public good and economies of scope in the production of the public good and private good (that is, payment services) by the public services) by the public payments system, Holthausen and Rochet (2005) show that the optimal pricing scheme involves a *subsidy* to using the public system (below-marginal cost pricing) in order to attract users, even when public funds to cover this subsidy are costly.

3.4 Access and tiering

Holthausen and Ronde (2002) examine access policies of national regulators who can restrict the access of banks to a net settlement system in a two-country model where the banking sector in each country is characterized by a monopoly bank with limited liability and coexisting net settlement and gross settlement systems. In such a banking system, banks with imperfect information inevitably make socially suboptimal decisions with respect to the use of the net settlement system. That is, banks allow their counterparties to use netting systems too often. The national regulator can increase national welfare by restricting access to the net settlement of the most risky banks. Banks whose access to net settlement is restricted resort to using gross settlement for funds in this model. However, this result is conditioned on regulators and banks having similar information about the riskiness of banks in the other country. The result is overturned if banks are assumed to have better information about their counterparties. Furthermore, this access policy is equivalent to one of mandating risk controls such as net debit caps for the most risky banks in net settlement systems. Hence, it is not an access issue per se, but one of mitigating the amount of risk in net settlement systems.

An interesting implication of Kahn and Roberds (2002) is that a desirable net settlement arrangement may feature a tiered structure. They consider an environment with limited enforcement that can be solved by posting costly collateral or using a costly and non-verifiable monitoring technology. Monitoring is socially preferable if its cost is low relative to collateral posting. When private agents incur lower monitoring costs than the central bank, it is efficient for them to undertake monitoring. However, since monitoring activity is non-verifiable, the incentives to monitor have to be provided by making "reliable" banks bear the burden of defaults by "unreliable" banks. As a result, it may be desirable to restrict the access of some agents to central bank settlement. To economize on the collateral requirements, reliable (or safe) banks are allowed to settle their transactions directly and have access to an uncollateralized, zero-interest intraday credit facility. *Potentially unreliable agents do not have access to central bank settlement, but instead settle through their settlement agent/monitor.* This hierarchy allows monitoring to substitute for collateral, and at the same time reduces the exposure of the taxpayer to credit risk.

Other literature on tiered payments systems examines the competitive consequences of the tiered arrangement. In this context, the clearing agent is an upstream provider of an essential input to indirect clearers. At the same time, clearing agents compete directly with the indirectly clearing banks in the retail market for payment services. Consequently, a clearing agent may face incentives to lever off its upstream position to gain a competitive advantage in the downstream market. Holthausen and Tapking (2004), and Tapking and Yang (forthcoming) demonstrate that this incentive exists within a tiered arrangement in payments systems. Lai, Chande, and O'Connor (2006) incorporate the extension of unsecured daylight overdrafts from a clearing agent to its indirect clearers. They show that the creation of credit risk by this unsecured credit facility can mitigate the clearing agent's incentive to raise the costs of indirect clearers.

3.5 Settlement rule: net versus gross

This subsection compares the costs and benefits associated with net and gross settlement systems. We will first consider an environment in which defaults of system participants are not a concern, and then introduce the possibility of defaults.

Without default

Angelini (1998) and Bech and Garratt (2003, 2006) analyze banks' incentives to postpone payments in an RTGS system in a two-bank model. Banks postpone payments in order to economize on the costly reserves needed to settle payments, and Angelini (1998) shows that this results in a dead-weight loss for the system as a whole, since the receiving bank faces more uncertainty over the timing of its inflows and needs to invest in more precautionary reserves. Since banks are ex ante identical, both can end up having to start the day with higher reserve balances to meet the payment obligations that can arise over the day.[14] Bech and Garratt (2003, 2006) formulate a normal-form game between the two banks and demonstrate that, while different central bank intraday policies (a collateralized credit regime or a priced credit regime) can change the set of equilibrium outcomes of the game, both regimes involve an equilibrium where banks postpone payments relative to the social optimum.

Kahn and Roberds (2001) use a neoclassical monetary model to study the liquidity constraint induced by an RTGS system. They show that, if intraday credit is not available, the implicit credit constraints imposed by RTGS distorts the intraday trading patterns.

[14] Unfortunately, Angelini fails to consider how the increase in the initial reserves of sending banks can mitigate the incentives to delay payment.

Temzelides and Williamson (2001) argue that when default is not a concern, an RTGS is Pareto dominated by a DNS system where settlement occurs with outside money at the end of each period. DNS, however, is still not an efficient arrangement. The efficient arrangement should involve having centralized payments and interperiod credit with no settlement imposed other than that households meet their lifetime budget constraints. The lesson is that, with perfect commitment, periodic settlement is not even an optimal arrangement. Therefore, an environment with default ruled out by assumption is not a desirable setting for analyzing different settlement rules.

With default

By adding an idiosyncratic investment return shock (aggregate uncertainty) where a bank can have high or low returns from their long-term investment, and consumers obtain an imperfectly informative signal about bank returns, Freixas and Parigi (1998) generate liquidity shortages and even *contagion* across locations, where consumption in one location is affected by signals at another location. Using this framework, which endogenizes liquidity and credit risks, the authors compare gross and net settlement. Unlike other studies, a trade-off exists between inefficient liquidation of productive assets under gross settlement and the inefficient survival of low-return banks under net settlement. The authors find that gross settlement is preferred if the proportion of inefficient banks is high, the opportunity cost of reserves is low, the cost of inefficient banks surviving is high, and the proportion of mobile consumers is small. Otherwise, net settlement dominates.

Freixas, Parigi, and Rochet (2000) extend the above set-up to allow for $N > 2$ banks and alternative travelling patterns (and hence patterns of payment flows) to investigate the systemic risk implications of a net settlement system. The authors identify two kinds of systemic risk. The first is a self-fulfilling banking crisis caused by coordination failure on the part of uninsured depositors. This is similar to the Diamond and Dybvig (1983) bank run, and can occur even when all banks in the economy are solvent and this is publicly known. Net settlement and the assumption that interbank payment obligations do not have priority over other claims on a failed bank are necessary conditions for this type of systemic event to occur.[15] The second type of systemic risk is that of contagion. Freixas, Parigi, and Rochet find that net settlement can, on the one hand, increase the resiliency of the system by spreading the losses among member banks, but at the same time reduce market discipline by

[15]The net settlement system considered here is said to lack "legal enforceability of netting," which guarantees the priority of clearing-house participants' claims so that netting cannot be unwound in the event of a default.

uninsured depositors, thus potentially leading to inefficiencies associated with excess forbearance. Contagion (the failure of one bank leading to the failure of other banks) can occur when payment flows between banks are sufficiently large. This risk is higher for certain patterns of payment flows, even assuming that these flows are balanced (or symmetric), so that each bank's multilateral net position is zero.

Kahn and Roberds (1998) examine the incentive of a representative bank to default on its payment obligations in a DNS system. The incentive of the bank to default is excessive relative to the social optimum, for two reasons: (i) default increases the priority of equity-holders relative to other claimants, and (ii) the social cost of bankruptcy is not borne by the defaulting bank. In contrast, under a gross settlement, default never occurs but social costs of reserves are incurred. However, given that bankruptcy costs are not too high, in the simplest case where bank asset quality is fixed and bank assets can always be liquidated at book value, net settlement dominates gross settlement. When the quality of bank assets is a choice variable, the cost of net settlement rises, due to negative effects on bank asset quality (that is, the bank overinvests in risk assets).

Kahn, McAndrews, and Roberds (1999) use a simple three-period model to show that the risk-reducing feature of an RTGS system is available only when delivery versus payment (DVP) is employed. In the absence of DVP, decentralized trading under gross settlement can lead to "mutually assured default" as a result of coordination failure. In this gridlock situation, all banks will default as long as they believe that other banks will also default. The possibility of payments gridlock creates a form of pure liquidity risk. Gridlock is not caused by any fundamental uncertainty about the value of banks' investments, but instead results from banks' distrust of their mutual creditworthiness. The threat of default by any one bank can cause a loss in priority of all interbank claims, thereby creating a situation where banks find default to be their best option. If banks anticipate this costly outcome, they will extend no intraday credit and no trade will occur. This coordination failure in settlement can be solved by posting a sufficiently high level of costly collateral so that incentives for strategic default are extinguished.

Kahn, McAndrews, and Roberds (1999) also illustrate that, in the case of symmetric trade flows, net settlement can eliminate gridlock equilibria without the need for posting costly collateral. Net settlement is effective because it de facto gives absolute priority to offsetting claims, since such claims are automatically discharged under net settlement rules. Net settlement can also avoid delays in trading that occur in gross settlement systems. Also, in cases where trading requires collateral, net settlement can economize on its use. Similar ideas are captured by Kahn and Roberds (2002) and Freeman and Hernandez-Verme (2004) in different settings. To summarize, netting offers an easy way of enforcing creditors' priority by replacing a difficult-to-enforce debt obligation with an inherently more enforceable one. By substituting participants' original payment obligation with an obligation to a centralized counterparty, a net settlement can allow for offsets that could not occur under a decentralized gross settlement system.

3.6 Intraday credit policies and risk controls

3.6.1 Pricing of intraday credit

Intraday credit policies are designed to mitigate the liquidity costs of gross settlement and the attendant risk of gridlock. If intraday liquidity is too costly, either due to an explicit interest rate or collateral requirements, then intraday credit has little effect on improving the efficiency of gross settlement systems. However, the provider of intraday liquidity can incur credit risk, and this has to be balanced against the efficiency gain of intraday credit provision. Various measures can be undertaken to reduce the amount of credit risk incurred by the liquidity provider. In addition to price (positive interest rate) and collateral requirements, there may be quantitative restrictions on the amount of liquidity a system participant is entitled to. The optimal design of intraday credit policy also has to take into consideration its distributional effect on all members of the system.

There are papers that take sides on this debate regarding whether intraday credit should be provided free or be costly.

On the one hand, some researchers argue that *intraday credit should be free* to reduce the phenomena described below.

(a) Distortions in consumption profile and distribution

Freeman (1996b) and Zhou (2000) construct a general-equilibrium model with a commodity market and an interbank settlement market. It is assumed that DVP is not feasible and settlement takes place a period after trading is conducted. Money is needed both for commodity trading and debt settlement. A settlement problem is introduced by assuming that the payment flows are uncertain and not fully synchronized, leading to a liquidity shortage. In particular, an intraday liquidity constraint arises due to agents' patterns of spatial separation. The shortage of liquidity may not be fully resolved by simple private intraday lending. An equilibrium with a binding liquidity constraint is suboptimal, since there is an uneven distribution of the allocation of consumption among ex ante identical traders. In this case, the central bank, with the ability to issue settlement money at zero cost, can improve welfare by providing liquidity. In the absence of credit risk, the interest rate on intraday credit should be set at zero, so that the liquidity constraint becomes non-binding. One interpretation is that efficiency requires the central bank to follow a sort of "Friedman rule" in terms of its intraday credit policy.

Kahn and Roberds (2001) develop a general-equilibrium model with DVP to study the RTGS-induced liquidity constraint. First, if intraday credit is not available, the implicit credit constraints imposed by RTGS distort the intraday trading patterns. Again, free intraday credit by the central bank allows payments system participants to overcome the credit constraints and thus achieve efficiency. They also show that efficiency cannot be achieved if the central bank makes credit available only against collateral. As well, efficiency cannot be attained by a policy of paying interest on reserves held overnight.

Freeman (1999) shows that, in the presence of aggregate default risk (in which a fraction of debtors will default with a positive probability), providing free liquidity for settlement is still optimal, even though the central bank suffers a loss when some borrowers of intraday credit do not repay. The loss is financed by printing new money. Therefore, with intraday credit, default results in an increase in the money stock and thus an increase in the price level. As a result, the free provision of intraday liquidity transforms default risk (borne by a subgroup of households) to inflation risk (borne by the whole economy). This default-induced inflation provides risk sharing by distributing default risk more broadly.

In a scenario where banks are members of a clearing house with the ability to enforce optimal reserve levels, and other non-deposit-taking intermediaries exist outside of the clearing house and are, therefore, not subject to the clearing-house-wide reserve requirement, McAndrews and Roberds (1998) demonstrate a role for the clearing house to provide free liquidity to clearing-house members. This maintains full insurance across depositors of its member banks, while member banks lend out available reserves to non-member intermediaries during a financial crisis, which occurs when there is a sufficiently large liquidity shock that there is an aggregate reserves shortfall.

(b) Payment delays

Angelini (1998) and Bech and Garatt (2003, 2006) consider the decision by banks on the timing of their payment transfers. When intraday credit is costly, and when banks are not co-operative, banks have the incentive to delay payments. This transfers the intraday credit cost to the receiving bank, generating a dead-weight loss to the payments system.

(c) Idle reserve holdings

Under a policy of a zero-interest reserve requirement, banks' excess reserves become idle balances. Kahn and Roberds (1998) study a bank's reserve-management problem when facing stochastic payment flows under different settlement arrangements. At the beginning of the day, a bank has to decide on its portfolio and chooses its amount of bank reserves. Because of the random nature of the payment flow, a bank may end up with a positive balance at the end of the day. The expected reserve is usually smaller in an RTGS system with free intraday credit than in one with costly liquidity: when free intraday credit is available, banks effectively need only sufficient liquidity to settle end-of-day net positions. Therefore, a policy of free credit reduces the chance of banks holding idle reserves at the end of the day.

On the other hand, some researchers argue that *intraday credit should be costly* for the following reasons.

(a) To cover monitoring and enforcement costs

Mills (2006) considers an environment in which agents have the choice to strategically default, and studies a policy of costly enforcement and pricing: the central bank invests in a costly enforcement technology that allows it to punish defaulters by confiscating some consumption goods. In this case, the optimal intraday interest rate may be positive, because of a requirement for the central bank to recover its costs of enforcement. This supports Rochet and Tirole's (1996) suggestion that the intraday interest rate be positive because monitoring and enforcement are costly.

(b) To reduce excessive risk taking

Kahn and Roberds (1998) show that, when default increases the priority of equity holders relative to other claimants, or when the social cost of bankruptcy is not borne by the defaulting bank, the provision of free intraday credit may induce banks to choose a more risky asset portfolio or even choose to default strategically. The presence of social costs to default that are not borne by a defaulting bank is another argument for a positive intraday rate, put forward by Lester (2005), who considers a setting with multiple settlements within a day.[16] When default occurs, the rest of the economy has to bear a social cost, which is increasing in the outstanding liabilities in excess of reserves. This cost captures (in reduced form) the potential contagion and the time and legal cost involved in recovering from a default. In this setting, when intraday credit is provided costlessly, banks do not have incentives to keep the optimal amount of reserves. As a result, intraday credit should be costly, to induce banks to hold more idle, liquid reserves, in order to reduce the size of potential default. Lester demonstrates that optimal settlement frequency and intraday credit policies should be jointly determined.

Notwithstanding the above pros and cons of free intraday liquidity provision, the debate can be further complicated by the presence of distributional effects. An intraday credit policy may also have distributional effects among heterogeneous agents. Most models suggest that a free intraday credit policy can achieve efficiency by making participants' liquidity constraints non-binding. Ghwee (2006) points out, however, that while free intraday credit can achieve Pareto optimality, it does not induce a Pareto improvement. This is because the injection of liquidity indirectly subsidizes buyers who purchase goods with debt at the expense of others, resulting in a redistribution of wealth among heterogeneous agents. Ghwee proposes a Pareto-improving policy that pays interest on reserves financed by inflation. Williamson (2006) also examines the redistribution effect of monetary injection. He builds a general-equilibrium monetary model with a rich clearing and settlement structure. Goods are purchased with credit, and these debts must be settled within the period. However, some debts are settled more quickly than others. If debt is settled quickly, then sales of goods can be used to finance other purchases by the household within the period. Prices depend on the payment instrument used. In particular, buyers have to pay a premium in a goods purchase if the corresponding debt does not clear quickly. This price premium depends on how tight the sellers' finance constraint is. Therefore, intraday liquidity provision can lead to relative-price changes.

3.6.2 Risk controls

Net settlement systems typically have in place risk controls to limit systemic risk. These risk controls can take the form of net debit caps, bilateral credit limits, collateral requirements, and loss-sharing rules that enhance system participants' incentives for mutual monitoring. Such risk controls can increase the cost of net settlement, and hence credit-risk reduction

[16] In general, intraday credit is provided for settlement (except in the extreme case of DNS), and default risk is not eliminated (except in the extreme case of RTGS).

needs to be balanced against these costs.

Collateral posting is desirable if it is costless, because it can provide free insurance against default. If, on the other hand, there is a positive opportunity cost of collateral, then requiring collateral can distort the allocation away from Pareto optimality. Collateral serves as an endogenous credit constraint (Mills 2006).

An alternative is to use a monitoring technology (for example, bank supervision) to distinguish a reliable participant from an unreliable participant (Kahn and Roberds 2002). When collateral is costly relative to monitoring, monitoring can help to reduce the collateral requirement for reliable participants. When there is monitoring, less resources are tied up in collateral. However, if monitoring effort is non-verifiable, incentives for monitors have to be provided by making them bear the burden of defaults by their clients. This results in a tiered or hierarchical settlement system, in which a monitor serves as a settlement agent for risky agents.

In Kahn and Roberds' (1998) model with strategic default by a representative bank, net debit caps can effectively lower the credit risks and the social costs of bankruptcy associated with net settlement. The optimal net debit cap, moreover, declines (becomes more restrictive) when endogenous asset-quality choice by the bank is allowed.

3.7 Summary of model implications

A clear message emerges from the literature that three frictions—–lack of double coincidence of wants, private information about trading histories, and limited commitment—–underlie the use of payments and settlement arrangements. The degree of information and commitment frictions will determine the optimal payment instrument, the optimal settlement frequency, and the optimal way to structure the payments system with regards to system provision provision and tiering. Being explicit about underlying frictions, as micro-founded models must do, helps us understand the role of central banks and public policy in mitigating the frictions. While tiering may have undesirable competitive consequences in a situation where a provider of an essential input to other firms also competes downstream with those firms, it may emerge as an optimal arrangement in which the incentives for peer monitoring are maintained.

Since payments and settlement arise in a world with limited information and commitment, a potential role arises for public policy with respect to mitigating those frictions. The agent responsible for the provision of the settlement asset, intraday liquidity, and finality (insurance) in the payments system should be able to commit to not overissue the settlement asset, to reabsorb liquidity provided intraday, to intervene during liquidity crises, and to have access to cheap liquidity production. Under some circumstances (when the optimal level of reserves maintained by banks is not self-enforcing), this agent should possess an enforcement technology to enforce optimal reserves. Although a private sector intermediary may be well able to perform the role of issuing a settlement asset and cheaply providing liquidity, a central bank may have several advantages with respect to commitment not to overissue the asset, to reabsorb the liquidity introduced by intraday needs, and to intervene in times of distress in the view of its responsibility for price and financial stability. Other reasons for central bank involvement in the payments system can include its neutrality with respect to other participants, economies of scope in the provision of settlement services and the pursuit of other public objectives, and its role as the government's banker.

With the focus on mitigating fundamental frictions, it is no surprise that researchers have used Type 2 or 3 models with micro foundations to address the issues of when payments and settlement are essential, and of central bank involvement in the provision of payments and settlement services. While a Type 1 model like Holthausen and Ronde (2002) has been useful for formalizing how privately optimal decisions regarding access to a system can be socially suboptimal, the issue of access should be studied in the context of a micro-founded model with clear-cut implications for social welfare, since access policies can shape the resulting structure of a payments system. For example, by restricting the access of the most risky banks, a market arises for the provision of settlement services by less risky banks to the most risky banks in the economy.

The optimal payments system design balances various risks and costs against each other. There are two categories of issues related to the design of payments system. The first involves the choice of real-time gross settlement versus deferred net settlement and the frequency of settlement within that mode. The second takes as given the settlement rule (net or gross), and determines optimal credit policies and risk controls within that framework.

Netting systems dominate gross systems when no default is considered, since the main risk of net settlement is the credit risk arising from unsettled payments. Even allowing for default, gross systems are often inefficient relative to net systems unless intraday credit is available at no or little cost. Since a key concern regarding net settlement is the creation of systemic risk through contagion, models with no default or with exogenous default scenarios do not capture the costs of net settlement very well. In fact, the incorporation of endogenous default risk, by raising the costs of net settlement, can overturn the above result, but how much credit risk is increased depends on the size and pattern (whether they are balanced or not) of payment flows, the proportion of risky/inefficient banks in the economy, and the social costs of default. By introducing credit risk to gross settlement and by increasing the cost of participation in net settlement systems, intraday credit policies and risk controls make gross settlement and net settlement systems more similar to each other with respect to the balance of risks and costs they entail.

The literature has addressed the issue of system design using a combination of Type 1, 2, and 3 models. Many studies use a reduced-form, partial-equilibrium (Type 1) framework to formalize the trade-offs between the settlement risk associated with net settlement and the liquidity cost (gridlock risk) associated with gross settlement. However, those studies cannot provide definitive answers regarding the optimal trade-off, since it depends on underlying fundamentals and frictions in the economy. Micro-founded (Type 2 and 3) models are necessary in this respect for meaningful welfare analysis and policy advice. Finally, it is worth bearing in mind that the various policy instruments and system design features should be jointly determined.

4. Modelling the Canadian Payments System

Research by Bank of Canada staff on specific Canadian payments and settlement systems, and on more general payments system issues, is important for developing the knowledge required to provide policy advice for payments and settlement system oversight.[17] Below is a description of recent research at the Bank of Canada. Most of these studies use simulator-based approaches, and the rest use reduced-form analysis (Type 1 models).

A group of researchers have focused on Canada's payments systems, particularly the Large Value Transfer System (LVTS). Arjani (2006) examines the trade-off between settlement delay and intraday liquidity using the simulation approach to test whether a complex queuing arrangement can lower liquidity costs and speed payments processing relative to a standard queuing algorithm. McVanel (2005) applies the simulation approach to assess the impact of unanticipated defaults in LVTS, by generating defaults in her model and then calculating individual survivors' capital positions following the defaults. McPhail and Vakos (2003) construct a reduced-form model of collateral demand by a representative bank in LVTS, and then test the model to see how well it can explain average holdings of collateral that banks

[17] An overview of the Bank of Canada's oversight role in Canadian payments systems is provided by Engert and Maclean (2006).

actually hold. Northcott (2002), on the other hand, examines the potential for contagion through linkages arising from the interaction of financial institutions in another Canadian payments system, the Automated Clearing Settlement System (ACSS), using a simulator of the ACSS. Caldwell (2006) constructs a game-theoretic model of the behaviour of two representative banks with respect to setting their bilateral credit limits in the face of anticipated defaults. Bilateral credit limits are an important feature of the Canadian LVTS.

Garcia and Gencay (2006) examine how the use of extreme value theory yields collateral requirements that are robust to extreme fluctuations in the market price of the asset used as collateral. Such considerations are important for securities settlement systems as well as other payments and settlement systems that require collateral to be posted by participants as a risk control. Lai, Chande, and O'Connor (2006), building on work from the industrial organization literature on vertically integrated firms, and on the literature regarding tiering in payments systems, investigate the impact of credit risk on the pricing of settlement services from the provision of uncollateralized overdrafts by a clearing agent to its indirect clearer in a tiered payments system.

Economic modelling of Canadian payments systems is still at an early stage. As we have argued in the introduction, economic models can potentially complement the existing "simulator-based" works by improving our understanding of the functioning of payments systems. Since most of the existing research is not tailor-made for studying the Canadian system, researchers may need to modify the existing models to better address questions specific to Canada. As we have discussed in section 2, model selection, in general, should be a function of the research question, and is subject to a trade-off between micro foundations, realism, and analytical tractability.

We next consider three examples to illustrate how that trade-off affects our model selection. First, suppose that a researcher aims to understand how LVTS participants choose their bilateral credit limits. In this case, one may need a game theoretical model with multiple banks, explicit credit limit choice, and realistic loss-sharing rules. Since we may need to capture many institutional details and we are interested in only the interactions within the settlement sector, we may construct a Type 1 model, taking as reference the game theoretical frameworks in Bech and Garratt (2003) or Caldwell (2006). However, if one suspects that bilateral credit limits may impact the efficiency of settlement and thus have an impact on the underlying trades in the economy, a Type 3 model, where the trading and banking sectors are explicitly modelled, would be a more appropriate framework. One may consider incorporating strategic behaviour into Lester's (2005) search-theoretic framework. Furthermore, if collateral is introduced in the model and a welfare evaluation depends on modelling the real cost of collateral, a micro-founded model (Type 2 or 3) would be necessary.

Second, suppose that a researcher wants to study the optimal degree of tiering in Canada. To address such a question, we would need a micro-founded model for payments system design, since we are ultimately interested in the "essentiality" of a payments system with a tiered structure and the welfare effects of such a structure. We can base our framework on the tiering model of Kahn and Roberds (2002) (a Type 2 model), to compare different institutional designs. We can also use a mechanism design approach (such as that used in Koeppl, Monnet, and Temzelides 2006) to examine whether tiering is an essential feature of a payments system. A Type 3 model can be used (for example, Lester, 2005) if researchers are interested in the effect of tiering on the bank-client relationship.

Third, suppose that a researcher aims to examine how Canada's banking structure may affect its settlement system. In this case, a Type 3 model is needed to examine how the banking industrial organization and strategic interactions among banks can affect the payment flows and incentives in the settlement sector. Useful starting points include Lester (2005), Freixas and Parigi (1998), and McAndrews and Roberds (1995). In these models, the structure of the banking industry determines the trading and payment flows and, accordingly, the optimal settlement policy.

5. Concluding Remarks

We have provided a framework for organizing diverse literature on payments systems into three broad classes. While there are several dimensions along which approaches to payments systems modelling differ, we choose to organize the literature with respect to three sectors that are explicitly modelled: the trading relationship, the bank-client relationship, and the interbank settlement relationship. Within each class of models, we provided benchmark examples and pointed out variants to the benchmark model. More importantly, we contrasted these model classes with respect to the questions they have been used to answer and are capable of answering, as well as their strengths and weaknesses. It is clear that one's choice of modelling tools depends not only on the question being asked but also on one's decision regarding the trade-off between tractability, realism, and micro foundation.

We have summarized the implications drawn from our literature review. In particular, we drew implications from the literature regarding the fundamental question of why payments and settlement are "essential" by considering the frictions (for example, trading, information, and commitment) that underlie the need for payments and settlement. We also summarized the policy implications across all three classes of models with respect to system provision, pricing, access, tiering, and system design, and related them to the model environment and assumptions.

We refer readers who are new to payments system research to a list of recommended reading in the appendix to this paper. Finally, we wish to emphasize that the payments system literature is still at an early stage, with many issues and questions not yet addressed. Some challenges and opportunities for future research are discussed below.

First, although some progress has been made to integrate an explicit banking sector into models of payments systems in which there is a separation between households and banks, and the bank-client relationship is endogenized, more work should be done to construct a model of payments systems where banking, payments, and settlement are essential. We envision this to be a model where banks serve the dual role of providing liquidity insurance and payments services to households. Within that framework, one can relate banking structure (an industrial organization concern) to payments system efficiency and stability; issues such as financial consolidation and the regulation of large complex financial institutions can also be studied. In addition, the issue of tiering in payments systems can be studied in such an integrated model. As mentioned, most economies' large-value payments systems exhibit some degree of the risks involved in a highly tiered payments system. How should a central bank respond to these risks? Put differently, what is the socially optimal degree of tiering in the payments hierarchy? Do central banks have a role in shaping the pyramid? Tiering can also create a concentration of payments within a small number of important correspondent banks. Under what circumstances does this create systemic risk?

Second, a challenge remains for the literature in terms of developing a unified framework for monetary and payments system policy analysis to examine the interactions between price stability and payments system policies. How does the design of payments systems affect the optimal monetary policy and the monetary transmission mechanism? Can payments theory explain why payments and settlement systems, as well as the conduct of monetary policy, vary across countries and over time?

Third, researchers have yet to build micro-founded models with multiple payments systems. Such models would be useful in examining issues related to the global environment, where there exist competing as well as complementary systems. This is pertinent, given the trend to internationalization of the financial services sector, with banks becoming global players and increasingly large international payment flows. Additionally, within a country, there are linkages across LVPS, smaller retail payments systems, and securities settlement systems. Is it sufficient to risk-proof each individual system? What are the key issues that can arise in an international arrangement for settling large-value payments? Will the consolidation of national payments systems lead to safer and more efficient settlement? On the one hand, this can diversify credit risk, but on the other hand it can concentrate payments in a few systems and introduce systemic risk. How should national authorities oversee such cross-border systems?

Finally, in all micro-founded models of payments systems, payment flows arise from transactions in the trade of goods. Financial markets are not explicitly modelled. In the real world, however, most of the large-value flows in payments systems arise from financial transactions. Is financial market microstructure relevant for the study of payments? Can payments system design affect payment flows related to financial portfolio adjustments?

The host of outstanding questions and issues, as well as modelling challenges, related to payments systems make it clear that this is an exciting and important area of research.

References

Angelini, P. 1998. "An Analysis of Competitive Externalities in Gross Settlement Systems." *Journal of Banking and Finance* 22(1): 1–18.

Arjani, N. 2006. "Examining the Trade-Off Between Settlement Delay and Intraday Liquidity in Canada's LVTS: A Simulation Approach." Bank of Canada Working Paper No. 2006-20.

Bech, M.L. and R. Garratt. 2003. "The Intraday Liquidity Management Game," *Journal of Economic Theory* 109(2): 198–219.

———. "Illiquidity in the Interbank Payment System Following Wide-Scale Disruptions." New York Fed Staff Reports No. 239.

Berentsen, A., G. Camera, and C.J. Waller. 2005. "Money, Credit and Banking." CESifo Working Paper Series No. 1617.

Berentsen, A. and C. Monnet. 2005. "Optimal Monetary Policy in a Channel System." University of Basel Working Paper.

Caldwell, G. 2006. "Anticipated Default in LVTS." Manuscript.

Cavalcanti, R. and N. Wallace. 1999. "A Model of Private Bank-Note Issue." *Review of Economic Dynamics* 2(1): 104–36.

Chakravorti, S. 2000. "Analysis of Systemic Risk in Multilateral Net Settlement Systems." *Journal of International Financial Markets, Institutions and Money* 10(1): 9–30.

Champ, B., B. Smith, and S. Williamson. 1996. "Currency Elasticity and Banking Panics: Theory and Evidence." *Canadian Journal of Economics* 29(4): 828–64.

Chapman, J.T.E. 2006. "Policy in International Payment Systems with a Timing Friction." University of Iowa Working Paper.

Committee on Payment and Settlement Systems (CPSS). 2003. "The Role of Central Bank Money in Payment Systems." Bank for International Settlements, CPSS Publication No. 55.

Diamond, D.W. and P.H. Dybvig. 1983: “Bank Runs, Deposit Insurance, and Liquidity.” *Journal of Political Economy* 91: 401–19.

Engert, W. and D. Maclean. 2006. “The Bank of Canada’s Role in the Oversight of Clearing and Settlement Systems.” *Bank of Canada Financial System Review* (June): 57–64.

Freeman, S. 1996a. “Clearinghouse Banks and Banknote Over-issue.” *Journal of Monetary Economics* 38: 101–15.

———. 1996b. “The Payments System, Liquidity, and Rediscounting.” *American Economic Review* 86 (5): 1126–38.

———. 1999. “Rediscounting under Aggregate Risk.” *Journal of Monetary Economics* 43: 197–216.

Freeman, S. and P. Hernandez-Verme. 2004. “Default and Fragility in the Payments System.” Texas A&M University Working Paper.

Freixas, X. and B. Parigi. 1998. “Contagion and Efficiency in Gross and Net Interbank Payment Systems.” *Journal of Financial Intermediation* 7(1): 3–31.

Freixas, X., B. Parigi, and J. Rochet. 2000. “Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank.” *Journal of Money, Credit and Banking* 32(3): 611–38.

Fujiki, H. 2006. “Institutions of Foreign Exchange Settlement in a Two-Country Model.” *Journal of Money Credit and Banking* 38(3): 697–720.

Furfine, C.H. and J. Stehm. 1998. “Analyzing Alternative Intraday Credit Policies in Real Time Gross Settlement Systems.” *Journal of Money, Credit, and Banking* 30(4): 832–48.

Garcia, A. and R. Gencay. 2006. “Valuation of Collateral for Extreme Market Events in Securities Settlement Systems.” Bank of Canada Working Paper No. 2006-17.

Ghwee, J. 2006. “Payments System and Monetary Policy Tools.” The University of Texas at Austin Working Paper.

Green, E.J. 1999. “Money and Debt in the Structure of Payments.” Federal Reserve Bank of Minneapolis *Quarterly Review* (Spring):13–29.

———. 2005. “What is the Role of a Central Bank in Payment Systems?” Pennsylvania State University Working Paper.

Green, E.J. and R.M. Todd. 2001. “Thoughts on the Fed’s Role in the Payment System.” Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 12–27.

He, P., L. Huang, and R. Wright. 2005. “Money and Banking in Search Equilibrium.” *International Economic Review* 46(2): 637–70.

Holthausen, C. and J.C. Rochet. 2003. “Efficient Pricing of Large Value Interbank Payment Systems.” European Central Bank (ECB) Working Paper No. 184.

———. 2005. “Incorporating a ‘Public Good Factor’ into the Pricing of Large-Value Payment Systems.” European Central Bank (ECB) Working Paper No. 507.

Holthausen, C. and T. Ronde. 2002. “Regulating Access to International Large-Value Payment Systems.” European Central Bank (ECB) Working Paper No. 22.

Holthausen, C. and J. Tapking. 2004. “Raising Rival’s Costs in the Securities Settlement Industry.” European Central Bank (ECB) Working Paper No. 376.

Kahn, C.M., J. McAndrews, and W. Roberds. 2003. “Settlement Risk under Gross and Net Settlement.” *Journal of Money, Credit and Banking* 35(4): 591–608.

Kahn, C.M. and W. Roberds. 1998. “Payment System Settlement and Bank Incentives.” *Review of Financial Studies* 11: 845–70.

———. 2001. “Real-Time Gross Settlement and the Costs of Immediacy.” *Journal of Monetary Economics* 47(2): 299–319.

———. 2002. “Payments Settlement under Limited Enforcement: Private versus Public Systems.” Federal Reserve Bank of Atlanta Working Paper No. 2002-33.

Kocherlakota, N.R. 1998. “Money Is Memory.” *Journal of Economic Theory* 81(2): 232–51.

Kocherlakota, N.R. and N. Wallace. 1998. “Incomplete Record-Keeping and Optimal Payment Arrangements,” *Journal of Economic Theory* 81: 272–89.

Koeppl, T., C. Monnet, and T. Temzelides. 2006. "A Dynamic Model of Settlement." European Central Bank Working Paper No. 604.

Lacker, J.M. 2005. "Payment Economics and the Role of the Central Banks." Speech given at the Bank of England Payments Conference, London, England. Available at http://www.richmondfed.org/news_and_speeches/presidents_speeches/index.cfm/2005/id=72/pdf=true.

Lacker, J.M. and J.A. Weinberg. 2003. "Payment Economics: Studying the Mechanics of Exchange" *Journal of Monetary Economics* 50: 381–87.

Lagos, R. and R. Wright. 2005. "A Unified Framework for Monetary Theory and Policy Analysis." *Journal of Political Economy* 113(3): 463–84.

Lai, A., N. Chande, and S. O'Connor. 2006. "Credit in a Tiered Payments System." Bank of Canada Working Paper No. 2006-36.

Lester, B. 2005. "A Model of Interbank Settlement." University of Pennsylvania Working Paper.

Martin, A. 2004. "Optimal Pricing of Intraday Liquidity." *Journal of Monetary Economics* 51(2): 401–24.

McAndrews, J.J. and W. Roberds. 1995. "Banks, Payments, and Coordination." *Journal of Financial Intermediation* 4(4): 305–27.

McPhail, K. and A. Vakos. 2003. "Excess Collateral in the LVTS: How Much Is Too Much?" Bank of Canada Working Paper No. 2003-36.

McVanel, D. 2005. "The Impact of Unanticipated Defaults in Canada's Large Value Transfer System." Bank of Canada Working Paper No. 2005-25.

Millard, S. and V. Saporta. 2005. "Central Banks and Payment Systems: Past, Present and Future." Bank of England Manuscript.

Mills, D. 2004. "Mechanism Design and the Role of Enforcement in Freeman's Model of Payments." *Review of Economic Dynamics* 7: 219–36.

Mills, D. 2006. "Alternative Central Bank Credit Policies for Liquidity Provision in a Model of Payments." *Journal of Monetary Economics* 53(7): 1593–1611.

Northcott, C.A. 2002. "Estimating Settlement Risk and the Potential for Contagion in Canada's ACSS." Bank of Canada Working Paper No. 2002-41.

Nosal, E. and G. Rocheteau. 2006. "The Economics of Payments." Federal Reserve Bank of Cleveland Policy Discussion Paper No. 14.

Rochet, J.C. and J. Tirole. 1996. "Controlling Risk in Payment Systems." *Journal of Money, Credit and Banking* 28(4): 832–62.

Sargent, T.J. and N. Wallace. 1982. "The Real-Bills Doctrine versus the Quantity Theory: A Reconsideration." *Journal of Political Economy* 90(6): 1212–36.

Tapking, J. and J. Yang. "Horizontal and Vertical Integration and Securities Trading and Settlement." *Journal of Money, Credit and Banking* Forthcoming.

Temzelides, T. and S. Williamson. 2001. "Payments Systems Design in Deterministic and Private Information Environments." *Journal of Economic Theory* 99: 297–326.

Townsend, R.M. 1989. "Currency and Credit In a Private Information Economy." *Journal of Political Economy* 97: 1323–44.

Williamson, S. 2006. "Transactions, Credit, and Central Banking in a Model of Segmented Markets." Manuscript.

Zhou, R. 2000. "Understanding Intraday Credit in Large-Value Payment Systems." Federal Reserve Bank of Chicago *Economic Perspectives*: 29–44.

Appendix A: Recommended Reading

For more details and to further your understanding, we suggest the following articles and books:

Non-Technical General Discussion

(i) Millard and Saporta (2005): development of payments systems and a central bank

(ii) Rochet and Tirole (1996): risk control in payments systems

(iii) Zhou (2000): intraday credit policy in a large-value settlement system

(iv) Green (2005): role of a central bank in a payments system

(v) Nosal and Rocheteau (2006): search theoretic models of payments systems

Theoretical Works

(i) Kahn and Roberds (1998): a benchmark Type 1 model studying the effects of settlement rules on banks' incentives

(ii) Freeman (1996b): a classic Type 2 model widely used in the literature

(iii) Kahn and Roberds (2002): a Type 2 model comparing the relative merits of privately operated, publicly operated, and tiering payments systems

(iv) Koeppl, Monnet, and Temzelides (2006): a Type 2 model using the mechanism design approach

(v) Lester (2005): a Type 3 model based on monetary search theory

(vi) Freixas and Parigi (1998): a Type 3 model based on the Diamond and Dybvig banking theory