14.462 Spring, 2020

Payment Systems: Problems and Issues to be Considered in Alternative Designs

(Lecture 8)

Robert M. Townsend

Elizabeth & James Killian Professor of Economics, MIT



Integrating Thread

Distinguish in what follows payments as a separate problem vs. payments connected to exchange and underlying economic environment

Walras without Obstacles: Pure Unit of Account in Commodities or Securities

- / commodities and a set Ω of households
- household *a* has an endowment $e_a \in R^l$
- utility function $U_a(f_a)$ where $f_a \in R^l$ and U_a is continuous, strictly increasing, and strictly quasi-concave
- allocation f is feasible if $\sum_{a\in\Omega}e_a\geq\sum_{a\in\Omega}f_a$
- An allocation f is called a Walras allocation if and only if there exists a p such that (p, f) is a competitive equilibrium with p · (f_a-e_a) ≤ 0 for all a
- An allocation f is a core allocation, if there does not exist a C ⊂ Ω, and an allocation g for all members of C such that ∑_{a∈C} e_a ≥ ∑_{a∈C} g_a and U_a(g_a) ≥ U_a(f_a) for every a ∈ C and the inequality strict for at least one a.

Application: Trade Fairs

Townsend (1990) – continued

- Consider a banking system where credit transfers are possible. That is, the obligation of the debtor (purchaser) can be eliminated by written transfer of a credit in the books of a banker. In effect, the debts of the banker can be reassigned.
- Further, bankers can create credit by writing entries in their books, even w/o the deposit of lawful currencies; a customer was simply granted an overdraft.
- Merchants used this banking transfer system in the great international trade fairs of Champagne. As noted earlier, such fairs had pre-specified days, first for introduction of goods, then for transactions in cloth, etc. There is little doubt that bankers played a key role in this sequence. Indeed, Usher (1943) believes that traders brought relatively few coins with them to the fair. Somehow potential sellers received credit from bankers for goods yet not sold, and this credit as buyers was transferred to other sellers relatively early on in the fair sequence



Ghost Currencies as Unit of Account

- A complete separation of a book-transfer payment system from coins or specie is hard to imagine. Again, however, it appears that medieval merchants came quite close at times.
- Supporting evidence is provided by the use at various dates and locations of pure units of account for record-keeping purposes, as noted by Cipolla (1956).
- Later, accounts were kept in circulating coins of various denominations to facilitate record-keeping. Subsequent devaluations changed the relative value of circulating coins, potentially destroying the record-keeping system. However, merchants in Milan after 1340, for example, continued to use older coins in their record-keeping systems with the understanding that they were pure fictional units of account



Trade Fails in Practice: Beneficial Impact of Expanded Fed Balance Sheet on Payments

- *As if an experiment with good, if unintended consequences
- Trade fails: previously high but now down
- Looking at payments more narrowly
 - Fed now makes available current and historical data on trades in US Treasury and other securities that fail to settle as scheduled
 - Substantial variation in the frequency of fails
 - surges in fails sometimes result from operational disruptions
 - But often reflect market participants' insufficient incentive to avoid failing



Source: Authors' calculations, based on data from the Federal Reserve Bank of New York. Note: The chart plots cumulative fails to deliver by month. http://libertystreeteconomics.newyorkfed.org/2014/09/measuring-settlement-fails.html

Why Do Treasury Trades Fail?

Settlement fails can occur for any of several reasons.

- Miscommunication. A buyer and seller may not have a common understanding of the terms of a trade, or one or the other may have failed to communicate settlement instructions to its custodian, or may have communicated incorrect instructions.
- System Failure. Well-known instance occurred on Thursday, Nov 21, 1985, with a computer outage at the Bank of New York . Also, after 9/11, when there was an initial surge in fails due to massive operational disruptions but insufficient incentive to resolve fails contributed to their persistence.
- Seller does not have the requisite securities in its commercial book-entry account. This is the most common reason for failing when fails are chronic, but it is usually avoided at other times by borrowing securities and delivering the borrowed securities. Otherwise the seller is not paid and is loosing implicit interest. There are other costs (capital charges, customer relations, back office staff costs, etc)
- Seller's failure to receive the securities in settlement of an unrelated purchase. This can lead to a "daisy chain" of cascading fails.

A Model of Actual Practice: Imposed Nash Game, Suggesting Current System Is Not Optimal

Bech & Garratt (2003) "The intraday liquidity management game"

Notes on Bech and Garratt

Mariano Spector

1 The Intraday Liquidity Management Game (2003)

- RTGS systems have the benefit of providing immediate finality to transactions, but can require large quantities of liquidity for banks to send out payments if they don't receive funds to offset them.
- This can create incentives for banks to postpone their payments as they wait for other banks to send them funds.
- Bech and Garratt analyze a dynamic game and show that the equilibrium can be inefficient, as in equilibrium banks may delay payments although the efficient outcome is to make the transfers immediately .

1.1 Model

In this model, payments are seemingly separated from underlying trades

- There are 2 banks and 2 periods (morning and afternoon of the same day)
- Banks start the day with a zero balance in their account.
- Each bank can receive in each period a request from its clients to make a payment (of \$1) to the other bank. The probability of receiving a payment request in the morning is p, and the probability of receiving a payment request in the afternoon is q (these events are independent across periods and banks). Payment requests received by each bank are private information.
- A bank that receives a payment request in the morning can choose to postpone it to the afternoon at a cost D (this represents the reputational costs of not fulfilling a client's request immediately). Payment requests received in the afternoon have to be fulfilled immediately.
- If the bank makes a payment without receiving funds to offset it, it has to pay F per dollar of overdraft as penalty.



Default and Need for Penalties to Implement Walras Optimum: Back to General Equilibrium Version of Payments

- Lessons from the Theory of Implementation
- In theoretical models, there are penalties which allow purchases before sales, or purchases in some markets and sales in others, which may or may not net to zero
- Dubey (1982): liquidity across markets and bankruptcy
- Moral of the Story: Penalties have to be sufficiently severe so as to achieve Walrasian outcome
- *But bear in mind some of the timing in Garrett et al is suppressed

- Define a market E which consists of
 - k different commodities.
 - Traders $i \in 1...N$
 - endowed with a vector $a^i \in R^k_+$ of commodities
 - continuous, concave, non-decreasing, utility function, which is strictly increasing in at least one variable.

- Assumption: For any commodity j, there exists at least two traders who are endowed with positive amounts of j and at least two traders prefer getting some of commodity j to nothing at all (some kind of competition)
- And, again, these could be in the space of securities and indirect utility function

- 1 trading post for each good $j \in 1, ..., k$
- agent *i* plays a static strategy $s_i \in S_i$ consisting of four vectors $\{p^i, q^i, \tilde{p}^i, \tilde{q}^i\}$
 - "if the price of commodity j is pⁱ_j or less, then I am willing to buy up to qⁱ_j units of j; if the price is pⁱ_j or more, I am willing to sell up to qⁱ_j units of j"
- Let $S = S^1 \times ... \times S^N$, an outcome function for agent *i* is a function $g^i : S \to R^k_+ \times R$ where $g^i(s) = (x, \beta)$ consists of a final bundle *x*, and net credit β that agent *i* receives when strategy profile *s* was played
 - if the total amount of unit of account an agent spends on good purchases exceeds the total amount he obtains from good sales, he will have a negative net credit

Supply and Demand Dubey (1982)



◆ロト ◆団ト ◆臣ト ◆臣ト 三臣 → の々⊙

The functions g^i can be characterized by the following two main properties:

- If the supply and demand curves intersect at a unique p^* : trade occurs between buyers with $p^i > p^*$ and sellers with $\tilde{p}^i < p^*$ and agents buy and sell the quantities they quote. If there is excess supply/demand the the marginal buyers and sellers (those who quote price p^*) are rationed in proportion to their demands and supplies.
- ② Buyers buy at the prices they quote. The highest buyer buys from the lowest seller. If he needs to buy more, he is serviced by the second seller and so on. If he does not satisfy supply of lowest seller, the lowest seller sells to the second highest buyer, etc.

 agents can borrow infinitely at a zero interest rate, but bear disutility for debt positions

•
$$P^i_{\lambda}(x,\beta) = u^i(x) + \lambda^i \min[0,\beta]$$

 if β < 0, expenses exceed revenues; if λⁱ = 0, then there does not exist maximizing strategy.

• define the game $\Gamma(E, \lambda)$ where the payoff function of player *i* is $_{\lambda}\Pi^{i}: S \to R$ where $_{\lambda}\Pi^{i}(s) = P_{\lambda}^{i}(g^{i}(s))$.

Competitive vs Nash Dubey (1982)

- focus our attention on active Nash equilibria
 - at each of the k trading posts, there are at least two active buyers and two active sellers

Lemma

At any active Nash equilibrium, all active buyers of commodity k quote the same price.

Lemma

In an active Nash equilibrium, $\beta^i = 0$ for every $i \in N$

Proof: At the end of any auction, $\sum_{i \in N} \beta^i = 0$ since any funds used to pay are given to the agent selling, therefore net credit cancels out. If there is a $\beta^i \neq 0$ then there must exist a $\beta^i > 0$, but then trader *I* could buy more of a commodity he likes and be strictly better off

Proof of Theorem Dubey (1982)

Theorem

For any E and $\lambda > 0$, the set of active Nash equilibrium of $\Gamma(E, \lambda)$ coincides with the set of competitive equilibria of E.

Any competitive equilibrium of E is an active Nash equilibrium of $\Gamma(E, \lambda)$

- Let $(\widehat{p}, \widehat{x}^1, ..., \widehat{x}^N)$ be a competitive equilibrium of the market E
- Let agent's *i* shadow prices on budget constraint be given by μ^i
- Choose αµⁱ < λⁱ for each i ∈ N (bankruptcy costs in units of utility; degree of freedom to choose numeraire with α)
- The below is a NE of $\Gamma(E, \lambda)$ yielding $(\widehat{x}^1, ..., \widehat{x}^N)$

$$p^{i} = \widehat{p}^{i} = rac{1}{lpha} \widehat{p}$$

 $\widetilde{q}^{i}_{j} = a^{i}_{j}$
 $q^{i}_{j} = \widehat{x}^{i}_{j}$

◆□ ▶ ◆□ ▶ ◆豆 ▶ ◆豆 ▶ ● □ ● ● ●

We omit the proof of the other direction

Another Take on the Problem, Linking Incomplete Contracts to Payments

Default and punishment in general equilibrium with incomplete markets

Dubey, Geanakoplos and Shubik (2005)

GENERAL EQUILIBRIUM THEORY has for the most part not made room for default. In the Arrow-Debreu model of general equilibrium with complete contingent markets (GE), and likewise in the general equilibrium model with incomplete markets (GEI), agents keep all their promises by assumption. More specifically, in the GE model, agents never promise to deliver more goods than they personally own. In the GEI model, the definition of equilibrium allows agents to promise more of some goods than they themselves have, provided they are sure to get the difference elsewhere. Agents there too must honor their commitments, though no longer exclusively out of their own endowments. Each agent can keep his promises because other agents keep their promises to him.

We build a model that explicitly allows for default, but is broad enough to incorporate conventional general equilibrium theory as a special case. We call the model $GE(R, \lambda, Q)$ because each asset j is defined by its promise R_j , the penalty rate λ_j , which determines the utility punishment for default on the promise, and the quantity restriction Q_j attendant on those who sell it. When

 λ and Q are set to infinity (or made sufficiently high), the model reduces to GEI.²

7. ENDOGENOUS DEFAULT PENALTIES

We turn to the dual of the problem in the last section, and show that when asset promises are exogenously restricted in \mathcal{A} , the market will endogenously choose intermediate default penalties in \mathcal{A}^* , even though higher and lower levels are available in \mathcal{A} .

We begin by asking how high the penalties *should* be, when promises are restricted.

7.1. The Economic Advantages of Intermediate Default Penalties with Incomplete Markets

There are four fundamental drawbacks to reducing the default penalties λ so far that some agents choose to default in at least some states in equilibrium: (i) creditors, rationally anticipating that they might not be repaid (on account of direct and indirect reasons), are less likely to lend; (ii) borrowers may not repay even in contingencies that have been foreseen, and even though they are able; (iii) imposing penalties is a deadweight loss; (iv) the default of unreliable agents imposes an externality on reliable agents who, because they cannot distinguish themselves from the unreliable agents, are forced to borrow on less favorable terms.

Despite myriad reasons why default is socially costly, the benefits from permitting some default often outweigh all of these costs. These benefits are basically twofold, and both stem from the fact that markets are incomplete to begin with. First, an agent who defaults on a promise is in effect tailoring the given security and substituting a new security that is closer to his own needs, at a cost of the default penalty. With incomplete markets one set of assets may lead to a socially more desirable outcome than another set. Second, since *each* agent may be tailoring the same given security to his special needs, one asset is in effect replaced by as many assets as there are agents, and so the dimension of the asset span is greatly enlarged. A larger asset span is likely to improve social welfare (although this gain must be weighed against the deadweight loss of the default penalties that are thereby incurred). In short, permitting default allows for a plethora of additional assets that do not have to be specified in advance.

A third benefit from allowing default, which is closely related to the first two, is that agents can go long and short in the same security, thereby doubling their asset span. We make use of this in the following example, which shows that the optimal default penalty is intermediate, even though it causes all the disadvantages (i)–(iv).

Massachusetts Institute of Technology

Liquidity Fails in Practice: A Systemic Risk

- The partial equilibrium payments view, again
- Kuussaari (1996): "Systemic Risk in the Finish Payment System: an Empirical Investigation", Bank of Finland
- On the payments system risk:
 - Risk that one clearing system participant's failure to settle causes other participants to default
 - Risk in clearing systems where payment messages are during day but funds transferred at the end of the day
 - Simulation of the effects of a systemic crisis in the Finish Payment system using actual payments system data.
 - ➢ If in the simulation a participant fails to settle, it is removed and algorithm is run again, removing potentially more entities until remaining banks are able to fulfill their obligations
 - > Thus, payments systems are linked directly to potential financial crisis
 - Though in the simulation these counterparty risk are judged to be small, they are real
- On the other hand, seems related to Chandrasekhar, Townsend and Xandri (2019)



Complexity in Practice, as in Computer Science: An Additional Obstacle for Decentralized Payments

- ✤Güntzer et al. (1998): "Efficient algorithms for the clearing interbank payments"
- The daily volume of the more than 50,000 payment orders currently to be processed in EAF-2 sums up to over DM 600 billion (approximately US\$ 353 billion)
- Complexity is a separate problem
 - Even when there is no uncertainty, as when both liquidity in accounts and trades are all submitted in advance, algorithms for clearing are computationally complex – NP hard
- Bank Clearing Problem: discrete optimization problem
 - > Objective function: clearing volume
 - Limiting resources: deposits of participants
 - Maximize number of completed trades
 - > Not all trades will be completed!
 - > Be mindful of limits within current centralized clearing systems
- Comparison to previous lectures
 - ➤ We have seen that scaling up is a problem due to validation, too
 - > But the problem here is not peculiar to DLT but rather includes traditional systems
 - > Is the solution hybrids with smaller numbers, as for off-chain?

Massachusetts Institute of Technology

Public Liquidity Is Not a Panacea

- Outside Central Bank money (coin, currency, other) as a way to store value for settlement is another way to achieve end-of-period clearing and settlement, and so can be seen as mitigating the role for penalties, if the central bank itself is trusted.
- But real-time-gross settlement (RTGS) systems are costly precisely in the requirement of setting aside large amounts of value in liquidity, instead of having this value invested or lent out for productive purposes.
- A time discount rate puts a wedge between money earned from sales and money used in purchases.
- Dynamics matter: Liquidity for payments comes from where?
- Thus RTGS settlements have in practice given rise to liquidity savings mechanisms, for example, contingent staggered settlement of purchases based on subsequent receipts.
- This returns us part of the way back toward where we started, and the problem of overspending and reneging. Ample central bank liquidity is no panacea.

The Cost of Liquidity: With Some Economics Coming from a Model

Townsend (1983) "Financial Structure and Economic Activity"

The basic model of the paper in terms of endowments, preferences, and technology, is presented in Section I. The model builds on Robert Lucas' version of the David Cass and Menahem Yaari (1966) circle as presented in my 1980 paper, modified here to allow for variable labor supply. Its key feature is the absence of double coincidence of wants for bilateral pairings, a feature which dates back to Knut Wicksell (1935), at least. In a Robinson Crusoe economy, in which households are completely isolated one from another. each household can consume at most the fruits of its own labor. This autarkic exchange regime is described in Section II. In a structure with spatially separated markets (essentially with bilateral pairings), a highly stylized asset, fiat money, partially overcomes the absence of double coincidence of wants. The decentralized, fiat money regime and its equilibrium are described in Section III. That regime is consistent with Clower's (1967) dictum that money buys goods and goods buy money, but goods do not buy goods. In a structure with centralized, Walrasian markets, another stylized asset, trade credit, delivers Pareto optimal allocations. The centralized, trade credit regime and its equilibrium are described in Section IV. In it there is a sense in which goods buy goods.

It is next established that the cost of market-produced commodities relative to home-produced commodities is infinite in autarky and is high in the decentralized, fiat money regime relative to the centralized. trade credit regime. The essential idea is that fiat money from the sale of home-produced commodities is held one period in the fiat money equilibrium (i.e., has unit velocity), whereas trade credit can be used for contemporary purchases in the trade credit equilibrium (i.e., has infinite velocity).² Thus, on the assumption that substitution effects dominate income effects, labor supply will increase, consumption of home-produced commodities will decrease, consumption of market-produced commodities will increase, trade will increase, and welfare will increase as one moves from autarky to the decentralized, fiat money regime to the centralized, trade credit regime. A formal analysis of



Application: Payments, Project Jasper, Canada as Actual DLT

- An experiment in private permissioned distributed ledgers allows for the exchange of central bank issued digital assets
- Could handle the high volume of Canada's large-value interbank transfer system
- Explored utilizing Ethereum, moved to R3's Corda platform to allow for improvements in settlement finality, scalability, and privacy
- The role of the Corda notary node is played by the Bank of Canada
- Netting promotes funding efficiency –a central queue within a DLT platform for payments
 - a participant's account gives permission for a bounded, specified amount of value to be placed into a queuing option. This can be changed, but not when the queuing algorithm is running; the participant is blocked as codes search over best transfers.

The central bank maintains a commitment to settle accounts but has risk exposure in doing so, collateral is posted by participants as part of DLT

Payments Canada (2017) "Project Jasper: A Canadian Experiment with Distributed Ledger Technology for Domestic Interbank Payments Settlement"

Process Step	Description	Initiator	Trigger Event	Pre- conditions	Post- conditions
Pledge	Pledge of CAD balance to the BoC	FI Participant	Submit DDR Obligation to BoC	DDR Obligation is correctly issued by Fl	DDR Obligation received by BoC node
Generate/ Fund	Generate CAD digital depository receipts	BoC	Acceptance of DDR Obligation from FI	DDR Obligation reviewed by BoC	DDR issued to requesting FI
Exchange - Atomic	Exchange DDR	Sending FI	Send DDR to Receiving FI	DDR available to support exchange	DDR consumed for Sending FI; DDR transferred to Receiving FI
Exchange – LSM	Add Payment – submit payment to queue	Sending FI	Send DDR LSM to Bank of Canada.	DDR is available to fund LSM exchanges	DDR LSM consumed on Sending FI -DDR LSM transferred to PC
	Netting – run LSM algorithm	LSM algorithm	LSM cycle time	DDR LSM Objects in LSM Queue available for aggregation	DDR LSM Objects reviewed with proper balance logic to conduct netting; triggers Exchange (LSM) Execute
	Execute – exchange net payments	PC	Exchange (LSM) Netting	Netting algorithm provides instruc- tions for atomic exchange	Instructions sent to relevant FIs to support multiple party atomic exchange
Redeem/ Archive	Request redemp- tion of DDR and archive	Sending FI	Submit DDR Obligation (Redeem) to BoC	DDR available for redemption of DDR Objects	DDR Object consumed on Sending FI -DDR Obligation (Redeem) sent to BoC -Copy of DDR Obligation (Redeem) remains with Sending Bank -DDR Object of requested amount sent to BoC where it is consumed (but saved)
Return	Return new net balance of DDR at BoC	BoC	Redeem accepted by BoC manually	Redeem DDR available	DDR Obligation (Redeem) returned to Sending FI for confirmation of archived DDR.



Garratt (2019) "An Application of Shapley Value Cost Allocation to Liquidity Savings Mechanisms"

This paper seeks to improve upon existing centralized netting queues by making two fundamental changes. First, instead of making decisions on how much liquidity to provide to the queue before netting arrangements are determined, banks receive take-it-or-leave-it offers that determine which of their payments will be settled as well as their share of the liquidity cost. This eliminates the need to solve a constrained integer programming problem. Second, rather than attempting to maximize the value or volume of payments settled in the queue, I propose using information regarding the instantaneous benefits and costs of participants in order to define a welfare measure for any set of netted payments. The full benefits of these two changes are realized through an application of the Shapley value cost allocation method, which ensures welfare maximizing netting proposals are always accepted.



Needed: Cross-Border Payments Systems with DLT

- Cross-border payments
 - ➤ Example: Remittances in fiat money in Southeast Asia have transfer fees currently at 7.1%. The high transfer fees are partly due to legacy technology in the formal sector and limited access to formal currency exchange markets

Bank of Canada and Monetary Authority of Singapore

Enabling cross-border high value transfer using distributed ledger technologies

Recognized priority for many central banks
 Hong Kong Monetary Authority
 Bank of England



Lightnet: Remittance Payment Rail for Money Transfer Operators in SEA

- Avoids direct transfers of fiat money, yet enables participants to efficiently conduct cross-border transactions
- MTOs who join have collateral placed securely and are granted credit lines for intraday transfers of fiat tokens
- Fiat tokens represent claims on fiat money
- Lightnet does the netting and liquidity management, the settlement layer, as in a central counterparty (CCP) in high value payments systems
- Payments rail
 - maintains an off-chain order book, groups and offset transactions before sending batched orders to Stellar at regular intervals for international transfer of fiat tokens. Deals with scaling problem

Fiat money does not cross borders



BANK FOR INTERNATIONAL SETTLEMENTS

Distributed ledger technology and large value payments: a global game approach

Stephen Morris (Princeton University)*
Hyun Song Shin (Bank for International Settlements)*

Keynes Lecture, University of Cambridge, 22 January 2019



Conference on "Cryptocurrencies and Blockchains", University of Chicago Becker Friedman Institute, 9 November 2018

* The views expressed here are those of the authors and not necessarily those of the Bank for International Settlements.

Where are we now?

Several central banks have experimented with DLT payment systems

- Cash-in-advance payment systems with digital tokens redeemable at central bank
 - Wholesale central bank digital currency (CBDC)
- Periodic netting arrangements to reduce credit needed for payments
- Central bank retains some role

Assessment so far

Technology works, but advantages over existing payment systems yet to be demonstrated



- ► How to ensure **confidentiality** of payments?
- How to overcome need for credit to finance payments?

Open, permissionless systems (eg, Bitcoin) have transactions visible to everyone, albeit with masked identities

Confidentiality of payments with oversight point to **private**, **hierarchical** DLT systems

- Private: banks and payment firms are voting nodes
- Hierarchical: central bank retains some role (eg, as notary)

Real time gross settlement (RTGS) systems have heavy credit needs

- \blacktriangleright Daily payments \approx 100 times the deposit balance held at central bank
- Banks rely on incoming payments to finance outgoing payments
- Bech and Garratt (JET 2003), Afonso and Shin (JMCB 2011)

Coordination problem arising from credit needs may swamp any technological refinements

Sources of funds in conventional domestic payment system

- 1. Balances maintained at the central bank
- 2. Borrowing from other banks through money markets
- 3. Credit extension from the central bank (eg, discount window or "daylight overdraft")
- 4. Incoming transfers from other banks

How to overcome **incentives to delay** when liquidity is scarce?

Who provides the credit to make the system work?