DEPFID
DEPARTMENT OF ECONOMIC POLICY, FINANCE AND DEVELOPMENT
UNIVERSITY OF SIENA

DEPFID WORKING PAPERS

Marco Galbiati and Simone Giansante

Emergence of networks in large value payment systems (LVPSs)

1 / 2010

DIPARTIMENTO DI POLITICA ECONOMICA, FINANZA E SVILUPPO
UNIVERSITÀ DI SIENA

DEPFID Working Papers - 1 / 2010

Emergence of networks in large value payment systems (LVPSs)

Marco Galbiati and Simone Giansante

Abstract

This paper develops and simulates a model of emergence of networks in an interbank, RTGS payment system. A number of banks, faced with random streams of payment orders, choose whether to link directly to the payment system, or to use a correspondent bank. Settling payments directly via the system imposes liquidity costs, which depend on the maximum liquidity overdraft incurred during the day. On the other hand, using a correspondent entails paying a flat fee, charged by the correspondent to recoup liquidity costs and to extract a profit. We specify a protocol whereby banks sequentially choose whether to link directly to the system or to become clients of other banks, thus generating a client-correspondent network. We calibrate our model on real data on the UK payment system, and we compare the networks it produces with i) the true client-correspondent network, ii) the outcomes of two 'dummy' benchmark models. The model is found to outperform the benchmarks. Its predicted networks reproduce some key features of the real UK network.

KEYWORDS: RTGS, network formation, tiering, correspondent bank, Nash bargaining.

JEL CLASSIFICATION: C7, G2.

ADDRESS FOR CORRESPONDENCE: Marco.Galbiati@bankofengland.co.uk. The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England.

ACKNOWLEDGEMENTS: The authors thank Mark Adams for contributions to an earlier version, and two anonymous referees, whose comments lead to substantial revisions. We are grateful to Prof. J Stiglitz for some insightful comments, and to Charles Kahn for very important suggestions. We also received sharp feedback from participants to: the 2008 Paris Conference on Computing in Economics and Finance (Society for Computational Economics), the 2009 NY Eastern Economic Association conference, and seminars at the City University (London), and at the CCFEA (University of Essex). The authors finally thank colleagues at the Bank of England for useful comments and support. The usual disclaimer applies.

Dipartimento di Politica Economica, Finanza e Sviluppo – Università di Siena

1 Introduction

Large value payment systems (LVPSs) are the electronic 'pipes' through which banks exchange payments with each other. Because cash payments are a tiny fraction of all transactions, and because all electronic transactions end up being between banks, it can be said that the whole economy of any advanced country 'runs through' its LVPS. In 2006, interbank payments in the UK's CHAPS system averaged £200 billion a day ($400 billion). The corresponding transactions in the US system (Fedwire) amount to about twice as much and, in the Euro area, volumes in the TARGET system largely exceed the 1 trillion dollars a day. Considering these staggering amounts, it natural that central banks and policy makers are interested in the smooth functioning of LVPSs, devoting substantial resources to their study, design and oversight.

These large aggregate flows are only part of the picture, as the structure of LVPSs differ drastically from country to country. In the UK for example, the main system has only 14 direct, 'first-tier' members, who settle payments on behalf of about 240 other institutions. At the other extreme, the US Fedwire system has a much less tiered structure: over 9500 banks, some of which are very small, link to the system directly and settle payments on their own behalf. A number of recent studies have charted the topology of payments over these networks in detail: Soramäki et al. (2007) look at the US Fedwire system; Becher et al. (2007) consider the UK CHAPS, Lublóy (2006) study the Hungarian VIBER, while Inaoka et al. (2004) look at the Japanese payment system BOJ-NET.

What lies behind these differences? Why do certain banks join a LVPS, while others make their payments via a first-tier correspondent? These questions are important because, first, the network structure of a payment system may affect the stability and efficiency of the system itself. Second, tiering implies that a share of interbank payments does not cross the official LVPS, settling instead on the books of the first tier banks.[1] Here, however, we do not attempt to clarify which structure is most desirable from a central bank's perspective. This paper concentrates instead on the following questions: what determines the structure of a payment network? To answer this, one must consider the incentives to join the first-tier of a LVPS, versus those to remain in the second-tier.

The literature on this is surprisingly scant. Lai *et al.* (2006) study, with a theoretical model, the contracts that may be signed between correspondent banks and their clients. The problem is interesting because correspondents may compete with clients on e.g. retail markets. Hence, this cannot be treated as a classic case of vertical integration. Lai *et al.* (2006), however, take the client-correspondent relationship as given; their model is not aimed at explaining how articulated networks form out of such relationships. Kahn et al (2005) also study tiered systems; however, their focus is on the social desirability of different structures, rather than on their emergence. A work with a similar objective is Harrison *et al.* (2005).

[1] This share can be large: it is estimated to be about 30% for the UK, or about £100 billion daily. For a discussion of risks involved in tiered payment systems, see Harrison et al. (2007).

As far as we are aware, there are only two works directly concerned with the topic of network formation in payment systems: Chapman *et al.* (2008) and Jackson et al. (2006). Chapman *et al.* (2008) present a model where correspondent banks 'monitor' the credit conditions of their clients, and can therefore ensure that these obtain funds for their risky projects. So, in Chapman *et al.* (2008), tiering is ultimately driven by credit risk and informational asymmetries. Although the authors present a fully-fledged model of network formation, its ingredients are completely different from ours: credit risk in their case, liquidity costs in ours.

In this sense, Jackson et al. (2006) are closer to our work: these authors do not have a model of emergence of networks, but they look at a bank's incentives to become direct member of a system, focussing on liquidity costs. Their findings suggest the existence of economies of scale in correspondent banking, generated by two effects: *internalization* of payments and *liquidity pooling*. Internalization refers to the fact that, when a bank acts as a correspondent, payments between its clients can be settled on the bank's own books ('on us'), at a zero liquidity cost. Liquidity pooling instead is a dynamic effect: by pooling uncorrelated payment requests from different clients, the liquidity need of a correspondent bank stabilizes, implying in turn that the costs of liquidity management are lowered. Internalization and liquidity pooling in CHAPS are estimated by Lasaosa et al. (2007) in a study which relates the degree of tiering to the liquidity needs of the system.

The dynamics of our paper are generated precisely by internalization and liquidity pooling. However, we innovate from Jackson et al (2006) in two senses: we describe these effects analytically and, more importantly, we plug these elements into a model of network formation. To anticipate, we set up a model where a number of banks face a random stream of payment requests. Banks can execute payments on their own, by borrowing liquidity from the central bank at a cost. Alternatively, they may become customers of other banks (correspondents), which would then execute payments on their behalf, thus relieving them of liquidity costs. However, correspondents charge their clients a fee to recoup costs and, possibly, to make a profit. Who becomes a correspondent, and who instead remains in the 'second tier' attaching to some correspondent, is endogenously determined by a dynamic process whereby banks 'shop around' for correspondents, who bid for clients. Making some assumptions on how corresponding services are priced, offered and accepted, we look at how the client - corespondent network evolves in time, converging to a stable state.

Our model is highly stylized. First, we model the timing of incoming and outgoing payments in the simplest possible way: pure randomness. That is, we abstract of any consideration about intraday liquidity management, an issue worth in itself a stand-alone paper (or more) -see e.g. Angelini (1988), Bech *et al.* (2003), who do not consider network formation, but payments choices.

Another, potentially more serious, simplification made in this paper is that we ignore credit risk issues that may emerge between correspondent banks and their clients. As we mentioned, these are elements that *can* influence the payment network structure. What we find interesting is that, even a parsimonious model like ours (with only liquidity costs) is able to replicate some features of a real network.

2 Model

The model has a population of N banks, sending payments to each other over a series of days. Banks can either be direct participants in the payment system, or they can hire a correspondent bank to execute payments on their behalf. If a bank participates directly, it needs to obtain liquidity from the central bank. If instead a bank decides to hire a correspondent, it only has to pay a price for the payment service. We look at how correspondent agreements evolve from day to day, leading to an equilibrium network.

This next two subsections describe a single day on a payment system as a sequence of events (payments). However, this is not the dynamic process we are most interested in. Our objective is the process of network formation, which unfolds on a longer time scale: *across* days, not *within* days. The single day model has therefore an instrumental role; it is condensed in equation 2 (below), and forms the basis for the model of Sections 2.3 and following -which is finally simulated to produce artificial networks.

2.1 Intraday payments

Banks are indexed by $i = 1, 2, .., N$. On any day, banks send to each other unit-size payments according to a 'payments matrix' $P = [p_{ij}]$. The various entries p_{ij} are integers representing the number of payments that i sends to j. We assume that incoming and outgoing payments are balanced:

$$\forall i, \quad \sum_j p_{ij} = \sum_j p_{ji} \equiv \frac{1}{2}\lambda_i \tag{1}$$

In real payment systems, payments need not be balanced on each single day. However, balancedness is an acceptable assumption when modelling an 'average' day. Indeed, this is the simplest way to avoid that some banks are cronically in deficit, and thus eventually disappear.

Payment volumes, and hence values, are fixed by P. However, we suppose that the sequence in which payments are made is random. The only assumption we make here is that the liquidity need of bank i, defined as the sum of payments sent *minus* payments received up to t, can be described as a symmetric random walk:

$$L_i(t) = \sum_{s \leq t} x(s)$$

where $\{x(s)\}$ is a sequence of independent r.v. equal to 1 or -1 with equal probability. The time elapsing between payments sent and received (ie the precise stochastic process generating payments) is unimportant here. To our purposes, equivalent formulations would be that payments arrival is a Poisson process (with intensity λ_i so that the average volume settled by i is λ_i), or that time is discrete and $L_i(t)$ is a random walk of length λ_j. Whichever interpretation we choose, this is irrelevant. Reason: the way we define costs, illustrated in the next section.

2.2 Liquidity costs and the 'pooling effect'

We assume that, if a bank *directly* participates in the payment system, it pays a cost related to its daily maximum liquidity need. In the UK, for example, banks participating to the CHAPS system have to pre-fund with the central bank their liquidity needs. If the central bank charges some function f for its overdraft, bank i's liquidity costs are $f(\max_t L_i(t))$, where t spans the lenght of the day. Depending on the (random) order according to which payments are made and received, the maximum overdraft varies from day to day. However, banks are supposed to be risk-neutral so they make decisions looking at:

$$C(\lambda_i) = E\left[f\left(\max_t L_i(t)\right)\right] \tag{2}$$

The Appendix I shows how $C(\lambda_i)$ can be computed. For certain specifications of the pricing function f, the expression becomes rather intricate. But, this is no concern: as we will fix the functional form of f and run simulations, all we need is an expression amenable to numerical computation -and Appendix I shows how to obtain it.

An important fact to note is that $C(.)$ is uniquely determined by λ_i, which is in turn determined by P. The following examples show how C looks like, for an f that we use in the simulations later on.

Example 1 *In the UK system CHAPS, banks obtain intraday liquidity from the Bank of England (BoE) at a zero interest rate, in exchange for collateral. Collateral is costly for for the banks to obtain in the first place. However, the so called 'double-duty' regime establishes that, to obtain intraday liquidity, banks may pledge with BoE some assets that they must hold* anyway *for prudential reasons. Hence, up to the value of these assets, liquidity costs are 'sunk' for CHAPS banks. For this system, a reasonable specification of the pricing function is therefore:*

$$f(x) = \begin{cases} 0 & \text{for } x < K \\ cx, \quad c > 0 & \text{for } x \geq K \end{cases}$$

where K is the amount of liquidity that can be obtained pledging double-duty assets. Figure 1 shows how the resulting expected costs $C(\lambda)$ depend on K. If $K = 0$, called 'type I', $C(.)$ is increasing, concave, asymptotically linear, with $C'(0) < 1$. If $K > 0$, called 'type II', $C(.)$ is 'S-shaped': first flat, then as in the above case. K has therefore the effect of 'flattening' C in the vicinity of $\lambda = 0$. For higher λs, costs are instead concave: this is the so-called 'pooling effect'*, which gives economies of scale in the payment activity.*

FIGURE 1

2.3 Correspondent banks and 'internalization effect'

Instead of participating directly in the payment system, a bank j may outsource its payments activity some other bank i. When this happens, bank i acts as the correspondent of bank j (which becomes client of bank i), and the following terms are agreed upon: i supports all liquidity costs deriving from j's payments, and in exchange j pays i a flat fee. A surplus is created by a correspondent agreement: first, some payments may be internalized. Second, there (may) be economies of scale from pooling payments, as shown by Fig. 1.

More precisely, suppose bank i is correspondent for a group of banks $S = \{i, j, ...\}$. In this case, bank i's liquidity costs are determined by the payments between S and the banks outside S. Instead, payments within S can be settled by changing entries of a book and require no liquidity. That is, bank is' cost will be equal to $C(\lambda_S)$, where

$$\lambda_S = \sum_{j \notin S} \sum_{i \in S} p_{ij} + \sum_{i \notin S} \sum_{j \in S} p_{ij} \qquad (3)$$

Note two that λ is sub-additive: given two groups A and B, $\lambda_{A \cup B} \leq \lambda_A + \lambda_B$. Indeed, if bank A has no payments from/to B, no payments can be internalized, so $\lambda_{A \cup B} = \lambda_A + \lambda_B$. But, if all payments made and received by B are towards A, then $\lambda_{A \cup B} = \lambda_A - \lambda_B$. In intermediate cases, $\lambda_A < \lambda_{A \cup B} < \lambda_A + \lambda_B$. *The subadditivity of* λ *is the so-called 'internalization effect'.*

Summing up: C is increasing but, due to internalization, adding a bank to a group S can either increase or decrease the costs of S's correspondent. As noted above (Figure 1) C is convex in a certain range so, even if $\lambda_{S \cup k} > \lambda_S$, it may still be $C(\lambda_{S \cup k}) < C(\lambda_S) + C(\lambda_k)$ i.e. a surplus may be realized by adding k to S.

2.4 Network formation

A 'network' is a partition of the N banks into groups, each with one correspondent. How do these groups form, i.e. how does the network evolve? We imagine that correspondent relationships are formed day after day, with banks accepting offers made according to the following protocol (index $t = 0, 1, 2...$ now refers to days).

1 At $t = 0$, all banks are self-settling;

2 at each $t > 0$, one randomly selected bank (say i) receives an offer from each other bank k; this is the fee k would charge i to become its correspondent;

3 i chooses the best (lowest) offer, becoming client of the best offerer;

4 when a bank i becomes a client of another bank, all its clients (if any) go back to self-settling. i must pay a penalty to its previous clients for breaching their contracts.

To clarify, the selected bank i receives offers from all correspondents and all clients (a client k makes an offer considering to leave its correspondent, to become itself a correspondent for the new group $\{i,k\}$). Of course i may also maintain its role; it will do so when the expected costs of doing so are lower than any other offer.[2]

The *offer*, or fee charged by the correspondent, is determined according to the Nash Bargaining Rule (NBR). In general terms, the NBR prescribes that, if parties a and b obtain a total profit ω by signing an agreement, they divide it in two shares x_a and x_b as follows:

$$\begin{array}{l} x_a = \frac{1}{2}\left(\omega + O_a - O_b\right) \\ x_b = \frac{1}{2}\left(\omega - O_a + O_b\right) \end{array} \qquad \text{(NBR)}$$

where O_i is what i receives if the agreement is not signed. In our story, party b (the client) pays a fee, so the offer is $-x_b$; the correspondent instead takes the remainder. It should be noted that these offers are 'myopic': banks do not consider that their partners might sign other contracts in the future.

The most well-known characterization of the NBR is that it prescribes the shares which would emerge from a repeated bargaining process, where two sides make each other offers and counter-offers in sequence. This interpretation may be not fit our model very well; however, a better (but less known) characterization is given by Young (1993). In Young's model, bargainers of a large population meet randomly pairwise, play the Nash demand game once[3], and then 'leave', to meet and play other individuals. And so on. Young shows that, if the individuals sample information on how the game is being played, and use such information to adapt their strategies, the population evolves to a population of Nash bargainers[4]. How does this fit our story? Imagine that the contracts have a fixed duration. Banks would meet, play the Nash demand game, stick to the resulting outcome for the duration of the contract, and then leave. After a while, they would have 'learned' to play according to the NBS. And that's where our story begins.

Example 2 *For simplicity, for a group A we write $C(A)$ instead of $C(\lambda_A)$. Suppose that k, a self settler with no clients, makes an offer to another similar self-settler i. If the offer is rejected, the parties' profit remain $-C(k)$ and $-C(i)$. If instead the offer is accepted, the total profit for both parties is $-C(\{i,k\})$. The NBR attributes to i a profit $\frac{1}{2}[-C(\{i,k\}) + C(k) - C(i)]$, i.e. i* is asked to pay $q_{ki} = \frac{1}{2}[C(\{i,k\}) - C(k) + C(i)]$. *This is the offer that k makes to i.*

Consider now the general case: i receives an offer from k. There can be two sub-cases: i is client of some w, or it is correspondent for group S.[5] In the first

[2] It's simple to see that no bank ever finds it convenient to go back to self-settling.

[3] In the Nash demand game, both parties simultaneously announce the shares they want, and get nothing if the shares are incompatible.

[4] A related result, for a matching a la Kandori, Mailath and Rob (1993), and for a learning behaviour, is given by Santamaria-Garcia J (2004).

[5] When i is a self-settlier, $S = \{i\}$.

sub-case, i's outside option is $O_i = -q_{wi}$. In the second sub-case, if i keeps its outside option, it bears a cost $C(S)$ but receives fees totalling $\sum_{r \in S \setminus i} q_{ir}$. So,

$$O_i = \begin{cases} -q_{wi} & \text{when } i \text{ is client of } w \\ O_i = -C(S) + \sum_{r \in S \setminus i} q_{ir} & \text{when } i \text{ is correspondent for } S \end{cases} \tag{4}$$

To determine the joint profits to i and k, recall that, if a correspondent i leaves its group, each of its clients goes back to self-settling. Hence, each 'abandoned' bank r suffers a loss of $C(r) - q_{ir}$. Bank i is liable for this, so its defection brings about a penalty of

$$X_i = \sum_{r \in S \setminus i} [C(r) - q_{ir}]$$

We suppose that i and its new correspondent k share this penalty. So when k, correspondent for group P, makes an offer to i, correspondent for group S, the profits for the new group are

$$\omega = -C(P \cup i) + \sum_{r \in P \setminus k} q_{ir} - X_i \tag{5}$$

Hence, the NBR prescribes that k charges i a fee equal to:

$$q_{ki} = \frac{1}{2}[-\omega + O_k - O_i] \ , \tag{6}$$

with ω defined in 5) and O_i defined in 4).

2.5 Dynamic properties

The abstract structure of the model is that of a 'coalitional game': we have a set of players N and, for each subset $S \subseteq N$, a payoff $C(\lambda_S)$ is given (Eqns. 3 and 2). To this coalition-form game, we attach a particular protocol (Section 2.4), specifying how coalitions form and dissolve. We don't pursue an abstract analysis of this game. However, the following fact provides theoretical ground for the simulations performed later on.

Lemma 1 *The network reaches a stable state in a finite number of steps.*

Proof. in Appendix II ■

Differently from the properties of the cost function C, which can be easily explored numerically, we *do need* an analytical proof of convergence for the network formation process. Indeed, checking convergence numerically would require exploring all conceivable paths -obviously an unfeasible task.

Lemma 1 ensures that no cycles are generated in our protocol; hence, an equilibrium *is* reached. *What* equilibrium then? The hub-and-spokes network (one bank acting as correspondent for all others), is trivially an equilibrium network.[6] However, it is easy to construct matrices P with two equilibria, both accessible from the same initial condition.[7] Because banks make decisions in a random order, one cannot speak of *the* equilibrium in general.

An analytical study of the statistical properties of these equilibria is beyond the scope of this work. Instead, we adopt a numerical approach. We calibrate the model on real data, and we run the protocol several hundred time using different 'seeds'. The resulting networks are then compared with i) a real network, and ii) the predictions yielded by a benchmark.

3 Testing the model

In this section we compare the outcomes predicted by our model with i) a real payments network and, to put things in perspective, ii) with the outcomes produced by a 'random benchmark'. The model will be judged on its ability to beat the benchmark in producing outputs close to reality.

As a reality target, we consider the network of the UK payment system CHAPS (we have a rich dataset on it). The testing exercise then requires making two somehow philosophical decisions:

a) Define a reasonable benchmark model, against which our model will compete. The benchmark needs to be smart enough (beating an easy looser provides a weak test), but at the same time not too artificial (a tautological model which exactly replicates reality is an unfair reason to dismiss other more parsimonious explanations).

b) Define a metric to evaluate 'closeness' between outputs, i.e. between the real network, the networks produced by our model, and those produced by the benchmark.

The next subsections illustrate our reality target, define benchmark and metric, presents our calibration strategy, and report the results.

3.1 Target

We aim our simulations at the UK payment network. This is composed of 14 correspondent banks[8], direct members of the system CHAPS, and 227 client banks, for a total of 241[9]. Correspondent banks are outside CHAPS, but are nevertheless within the payments network -which is indeed defined as the client-correspondent network. Figure *2* shows our target network:

FIGURE 2

[6] A lone correspondent internalizes all payments, and so it incurs zero costs; as a consequence, its fees cannot be undercut.

[7] If two groups have many within-group payments, and few cross-group payments, they are somewhat "far" from each another, and they can co-exist in equilibrium.

[8] Plus the Bank of England, which however we subtract from the picture. BoE does not offer correspondent services as commercial banks, nor faces liquidity costs.

[9] This was in 2003, at the time of the data we look at.

It can be noted that a few banks have more than one correspondent; this is a feature that our model will *not* be able to replicate -our protocol assumes that each bank has at most one correspondent. On the other hand, it should be noted that banks with 2 correspondents typically use only one, with the other link being kept as a back up measure.

Table 1 shows some statistics of the payments network:

Table 1: Some statistics of the UK payments network

Correspondents	*Internal. payments*	*Gini payments*	*Gini customers*
14	33%	0.61	0.69

Column 2 records the percentage of internalized payments, ie payments made within groups of clients. Central Banks regard the percentage of internalized payments as a key feature of a payment system. The reason, we mentioned, is that these payments settle outside the official systems. Technically, internalized payments settle in 'commercial bank money' (on the books of commercial banks), and they are often accompanied by extensions of credit. Hence, they may originate risks, from which their relevance for central banks.

Column 3 record the Gini concentration index, computed on the volumes of payment settled by each correspondent. This shows that the UK system is fairly concentrated, in terms of payments activity. Similar conclusions can be drawn from column 4, reporting the Gini index, computed on the number of clients of each correspondent bank. Concentration of the correspondent business is also an issue of interest for Central Banks and policy makers: higher concentration implies higher liquidity efficiency, but may also imply that risks (e.g. operational risks) are more concentrated.

3.2 Benchmark

As a benchmark to compete against, we choose two models of random network formation, with and without preferential attachment. Given the structure of our 'target' network, a classic (e.g. Erdős–Rény) random graph would be too easy a benchmark to outperform. Indeed, a serious benchmark for our case should produce random graphs with only two tiers, and with low connectivity (the correspondents and their groups form disconnected subgraphs). To make the benchmark more interesting we also want that, on average, it produces the correct number of correspondents (14).

Our benchmark thus is the following: each of the 241 banks is picked in random sequence. With probability 14/241, the chosen bank is made a correspondent; with probability $1 - 14/241$, it is made a client and is attached to an existing correspondent randomly -with or without preferential attachment, depending on the specification.

Such benchmark clearly abstract from the 'geography' of payments, which is instead key in our model.

3.3 Distance (and other success measures)

There are a number of definitions of distance between networks. We are not experts on this, but to our understanding these broadly fall under three types.

Type I - metrics defined on algebraic properties of the adjacency matrices (spectral distances, entropy-based distances). We didn't find any of these particularly meaningful in our context. Also, these metrics present drawbacks when applied to disconnected graphs like ours.

Type II - metrics based on 'counts' of the links that need be moved to transform one graph into another. These metrics have a more intuitive meaning and can be applied to all contexts.

Type III - metrics designed to measure the distance between *partitions* of a set, rather than between *networks* in general. These measures are very suitable here: our model has some features of a coalition formation game, and the networks it produces do represent partitions (with the additional twist that one bank in each group has a special role -the correspondent).

Our first measure is of Type II. It is defined as the fraction of links that need be changed to reach the target network, i.e. the 'real' system. Recall that banks have exactly one correspondent; so a network here can be represented by a $N \times 1$ vector κ, with k_i being the i's correspondent. Our measure then is:

$$D1\left(\kappa, \kappa'\right) = \sum_i \frac{I_{\neq 0}\left(\kappa_i - \kappa_i'\right)}{N} \tag{7}$$

where $I_{\neq 0}(x) = 1$ iff $x \neq 0$. A **low** value of $D1$ means that κ and κ' are close.

The second metric we use belongs to the family of distances between partitions. It is a modified version of the Rand index (Rand (1971)), which measures similarity in the way nodes are clustered. To define it, let s be the number of pairs of banks which are under the same correspondent in network κ, and which are under the same correspondent in network κ': $s = \left|\left\{(i, j) : k(i) = k(j) \wedge k'(i) = k'(j)\right\}\right|$. In a sense, this is the number of successes in predicting that i and j share a correspondent. Define then $z = \left|\left\{(i, j) : k(i) = k(j)\right\}\right|$ -this is the number of banks with a common correspondent in network k. Our second measure is:

$$D2\left(\kappa, \kappa'\right) = 1 - \frac{s}{z} \tag{8}$$

Hence, κ' is 'close' to κ, if is able to put in the same group banks which *are* in the same group in κ[10]. A **high** value of $D2$ means that κ and κ' are close.

A third, simplest measure, gauges how successful the model is at detecting real CHAPS correspondents. The rationale behind this measure is that correspondent banks are 'special' nodes in the network, so they deserve special

[10] This measure can be made more demanding, by counting as success only a correctly identified pair, but only when the correspondent too is correctly identified. To anticipate, if we use such a measure, our model comes out even more as a winner.

attention. For any network κ, be CS_k the set of its correspondents. Denoting the real network by κ^*, the success rate of κ is defined as

$$\sigma(\kappa) = \frac{|CS_k \cap CS_{k^*}|}{|CS_k|} \tag{9}$$

That is: $\sigma(\kappa)$ is the fraction of correctly identified correspondents, out of all predicted correspondents. A **high** value of σ means that κ is successful in mimicking κ^*.

Finally, we also look at the ability of the model to match the properties of the UK system reported in Table 1. As explained there, these are *the* key statistics in a payment system, from a policy point of view.

3.4 Calibration

The model is very parsimonious, so calibration is relatively simple: it only requires specifying two inputs: the matrix of payments P (eq. 1), and the cost function f (eq. 2).

For function f, we adopt the specification of Example 1. As commented there, threshold K represents a bank's regulatory prudential assets, and thus is a function of a bank's size. We then posit $K = T\lambda_i,\ T > 0$. That is: we consider K a linear function of a bank's size, and we proxy a bank's size by its payments activity. Comparing data on collateral holdings and payment activity for UK banks (both available to the BoE), one ends up with an estimate of T in the range $0.1 - 0.2$.[11] We will run simulations for $T = 0$ and $T = 0.125$. Although probably at the time of the data (2003) T was above zero, a value of $T = 0$ is an interesting case, as it could represent a tightening in liquidity regulation (abolition of the 'double duty').

For the matrix P we proceed as follows. BoE has a complete record of all transactions taking place on the CHAPS -which provides the backbone of matrix P. However, CHAPS transactions are by definition those between correspondent banks. Unfortunately, when a transaction is made on behalf of a client, the underlying payer and payee are not recorded in CHAPS. To reconstruct the full matrix P we then use the Bank of England 2003 CHAPS traffic survey dataset (also used in Becher et al (2007)). This survey samples five days of the payments executed on CHAPS, recording both the ultimate payer and payee banks, and the correspondents used. Crucially, the survey also asks correspondent banks to report the percentage of internalized payments. Allocating these payments between banks who use the same correspondent, in proportion to their outgoing payments over CHAPS as revealed by the survey, we obtain a P to use in the simulations.

[11] The range arrives at 0.3 if one considers only UK banks in CHAPS. However, non-UK banks in CHAPS are subject to a different liquidity regime from the 'double duty'. For them α is probably close to zero.

3.5 Results

All the results presented here are obtained running 400 simulations of the model (200 for $T = 0$, and 200 for $T = 0.125$), and 400 of the benchmark (200 for the preferential attachment version, and 200 for the other).

3.5.1 Matching basic features of the real system

As a first test, we look at the success rate σ: how good the model is at detecting real CHAPS correspondents. Figure 4 a shows that the model achieves a much higher success rate than the benchmarks.[12] To put things in perspective, Figure 5 shows the corresponding cumulative distribution functions. It turns out that the 'worst' 5% networks generated by the model outperform 98% of the benchmark-generated competitors.

Figures 6 is about the percentage of internalized payments. In the real system, internalized payments are 33%, so the model is doing clearly better than the benchmarks (which produce too low underestimates). At $T = 0.125$, the model gives an almost unbiased prediction -although its modal prediction falls somehow short of the target.

Finally, Figure 7 refers to the Gini concentration index on volumes. At both specifications of T, the model gets only moderately upwards biased predictions of the true 0.61 value (0.63 on average for $T = 0$, and 0.65 on average for $T = 0.125$). Instead, both benchmarks clearly underpredict the Gini index: the one shown, with preferential attachment, averages 0.54 (while the other benchmark does even worse, predicting 0.39 on average and with a similar dispersion). Essentially the same conclusions are arrived at looking at Gini indices on the number of clients.

3.5.2 Distance measures

As a final test, we turn at the measures of Eqns. 7) and 8). We recall that $D1$ a distance, to be minimized, while $D2$ is a closeness index, to be maximized.

Figure 8 clearly illustrates that the model largely outperforms the preferential-attachment benchmark on $D1$ (the other benchmark is not shown as it scores even higher distances). In particular, $T = 0.125$ appears to be a better choice of parameter.

A similar message transpires from Figure 9: the nets produced by the model score higher in $D2$ than their competitors (again, we don't show the non-preferential attachment benchmark, as it performs slightly worse). As for $D1$, the model produces best results at $T = 0.125$.

[12] Both benchmarks (preferential e non-preferential) perform equally in terms of success rate. Indeed, they differ in how clients are attached to correspondents, not in how correspondents are chosen.

4 Conclusions

This paper formally modelled the ‘netting’ and ‘liquidity pooling’ effects, exploring how these can shape the client-correspondent network of a payment system. The model is extremely parsimonious, requiring essentially two inputs: a matrix of payments (P), and a liquidity price function (f). Still, when a parametrization is performed using data on the main UK payment system, the model produces networks which match a number of important features observed in reality. In this, the model is shown to outperform two benchmarks models of random network formation.

5 Appendix I - computing Eq. 2)

Recell that $L_i(.)$ is a symmetric random walk of length λ (we drop the index i, unnecessary here). So,

$$\begin{aligned} C(\lambda) &= E\left[f\left(\max_t L(t)\right)\right] \\ &= \sum_z f(z)\, p(z,\lambda) \end{aligned} \tag{10}$$

where $p(z,\lambda) = prob\left[\max_{t=0..\lambda} L(t) = z\right]$, and z runs over all possible values that $L_i(t)$ may take as $t = 0...\lambda$. That is, summation runs from $z = -\lambda/2$ to $z = \lambda/2$ (recall that $\lambda = 2\sum_j p_{ij}$, so λ is even).

Now, the distribution function of the maximum of a random walk is well known (it can be obtained using the reflection principle):

$$p(z,\lambda) = \left[\binom{\lambda}{r+1} + \binom{\lambda}{r}\right] p^{\lambda}$$

This can be plugged into eq . 10) so, once the functional form of f is specified, costs are an expression which can be immediately computed.

As incidentally stated at the end of Sect.2.3, $C(.)$ is increasing for all increasing f. Given what we said, the proof is rather intuitive: increasing λ 'lengthens' the random walk L. So the pdf of its maxima is skewed towards higher values. As f is increasing, the effect on the sum C turns out to be positive.

Similarly, the existence of a 'pooling effect' (convexity of C) that appears with a linear f and discussed in Sect. 2.2, can be easily proven by differentiating twice eq. 10) with respect to λ.

However, in the main text we do assign a specific form to f, and compute C numerically. So, analytical proofs about the cost function are somehow redundant.

6 Appendix II - proof of lemma 1

For a given network $\Xi = \left\{S^1, S^2, S^3...\right\}$, consider the network's total costs $TC = \Sigma_i C\left(S^i\right)$. We prove that TC is a Lyapunov function for the protocol: every time the network changes, TC decreases. So TC eventually reaches a minimum, at which point the network stops changing.[13] Behind this, the constant application of the NBR to determine both offers and penalties: a new correspondent relationship is established only if some 'extra profit' is realized.

Suppose then k accepts i's offer to join S; there are two cases: a) k is correspondent (for say P), b) k is client of $w \neq k$.

[13] With a finite number of banks and networks, TC takes on a finite number of values.

Case a) If k accepts i's offer it must be $\omega > O_i + O_k$ (NBR). That is:

$$\begin{aligned}
-C\left(S \cup k\right) + \sum_{r \in S \backslash i} q_{ir} - X_k &> \left(-C(S) + \sum_{r \in S \backslash i} q_{ir}\right) + \left(\sum_{r \in P \backslash k} q_{kr} - C(P)\right) \Rightarrow \\
C\left(S \cup k\right) - \sum_{r \in S \backslash i} q_{ir} + \sum_{r \in P \backslash k} \left[C\left(r\right) - q_{kr}\right] &< C(S) - \sum_{r \in S \backslash i} q_{ir} + C(P) - \sum_{r \in P \backslash k} q_{kr} \Rightarrow \\
C\left(S \cup k\right) + \sum_{r \in P \backslash k} C\left(r\right) &< C(S) + C(P)
\end{aligned}$$

On the l.h.s. there are the costs of all banks affected by k's decision[14] when k joins S and P is disbanded. On the r.h.s., the costs that would obtain otherwise, with S and P unchanged. Thus, if k joins S, TC falls. The same inequality can also be obtained from $q_{ik} < -O_k$ (k prefers i's offer to its own profits as correspondent).

Case b) Suppose k is a client of $w \neq k$, correspondent for some P. If k accepts i's offer, this must be more convenient than w's offer:

$$\begin{aligned}
d_{ik} &< d_{wk} \Rightarrow \\
\frac{1}{2}[-\omega\left(i,k\right) + O_i - O_k] &< \frac{1}{2}[-\omega\left(w,k\right) + O_w - O_k] \Rightarrow \\
-\omega\left(i,k\right) + O_i &< -\omega\left(w,k\right) + O_w \Rightarrow \\
\left(C\left(S \cup k\right) - \sum_{r \in S \backslash i} q_{ir}\right) + \left(-C\left(S\right) + \sum_{r \in S \backslash i} q_{ir}\right) &< \\
\left(C\left(P \cup k\right) - \sum_{r \in P \backslash w} q_{ir}\right) + \left(-C\left(P\right) + \sum_{r \in P \backslash w} q_{ir}\right) &\Rightarrow \\
C\left(S \cup k\right) - C\left(S\right) &< C\left(P \cup k\right) - C\left(P\right) \Rightarrow \\
C\left(S \cup k\right) + C\left(P\right) &< C\left(P \cup k\right) + C\left(S\right)
\end{aligned}$$

On the l.h.s. there are the costs when k joins S instead of P; on the r.h.s., the costs that obtain otherwise. Thus again, if k joins S, TC fall. The same inequality would obtain from $\omega > O_i + O_k$ i.e. for k to join i, the surplus to share must exceed the sum of the outside options.

[14] Bank K's acceptance affects only the costs of i, k, and of the banks in P.

References

[1] Armantier, O, Arnold J and McAndrews J (2008), "Changes in the Timing Distribution of Fedwire Funds Transfers", forthcoming on the FRBNY Economic Policy Review.

[2] Angelini, P (1998), "An Analysis of Competitive Externalities in Gross Settlement Systems", Journal of Banking and Finance n. 22, pages 1-18.

[3] Bech, M L and Garratt, R (2003), "The intraday liquidity management game", Journal of Economic Theory, Vol. 109(2), pages 198-219.

[4] Becher, Soramaki and Millard (2007), "The topology of CHAPS payments" - forth. Bank of England wp.

[5] Becher C., Galbiati M., Tudela M. (2008), "The Timing and Funding of CHAPS Sterling Payments", Economic Policy Review, Vol. 14, No. 2, September 2008.

[6] Chapman J, Chiu J, Molico M, "A model of settlement networks", mimeo.

[7] Galbiati M and Soramaki K (2007), "An agent-based model of a payment system", Band of England wp No 352.

[8] Harrison, S, Lasaosa, A and Tudela, M (2005), 'Tiering in UK payment systems', Bank of England Financial Stability Review, June, pages 63-72.

[9] Inaoka H., Ninomiya T., Taniguchi K., Shimizu T., Takayasu H. (2004) "Fractal Network derived from banking transactions - An analysis of network structures formed by financial institutions", Bank of Japan Working Paper Series, No. 04-E-04, April 2004.

[10] Jackson, J and Manning, M (2006) 'Central bank intra-day collateral policy and implications for tiering in RTGS payment systems', mimeo.

[11] Kahn, C M and Roberds M (2005) 'Payments settlement: tiering in private and public systems', University of Illinois and Federal Reserve Bank of Atlanta, mimeo.

[12] Lai A, Chande N, O'Connor S, "Credit in a tiered payment system", Bank of Canada working paper, 2006-36.

[13] Lasaosa A, Tudela M, "Risks and efficiency gains of a tiered structure in large-value payments: a simulation approach", BoE forthcoming WP.

[14] Lublóy Á (2006), "Topology of the Hungarian large-value transfer system", Magyar Nemzeti Bank, Occasional Papers 57, 06.

[15] McAndrews, J. and Rajan, S. (2000), 'The Timing and Funding of Fedwire Funds Transfers', Federal Resever Bank of New York Economic Policy Review 6 (2), 17-32.

[16] Norris J R. (1997), "Markov Chains", Cambridge Series in Statistical and Probabilistic Mathematics, CUP.

[17] Rand W M (1971), "Objective criteria for the evaluation of clustering methods". Journal of the American Statistical Association 66: 846–850.

[18] Santamaria-Garcia J (2004), "Equilibrium Selection In The Nash Demand Game. An Evolutionary Approach", IVIE wp 2004-34

[19] Soramäki, Kimmo, M.L. Bech, J. Arnold, R.J. Glass and W.E. Beyeler (2008). "The Topology of Interbank Payment Flows". Physica A. Vol. 340, pp. 380-394.

[20] Harrison S, Lasaosa A, and Tudela M (2005), "Tiering in UK Payment Systems: Credit Risk Implications". Bank of England Financial Stability Review Issue 19 pp. 63-70.

[21] Young H P (1993), "An Evolutionary Model of Bargaining," Journal of Economic Theory, 59 (1993), 145-168.

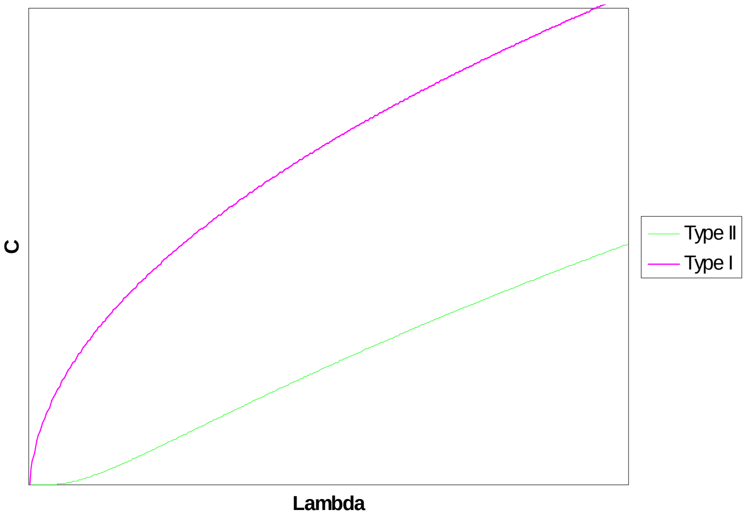

Figure 1: Two cost Types: $K = 0$ and $K > 0$

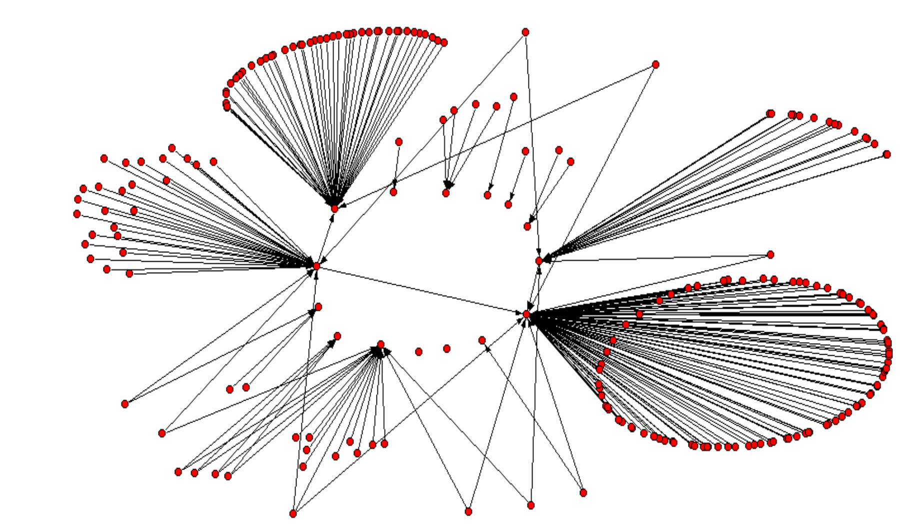

Figure 2: UK clients-correspondents payments network

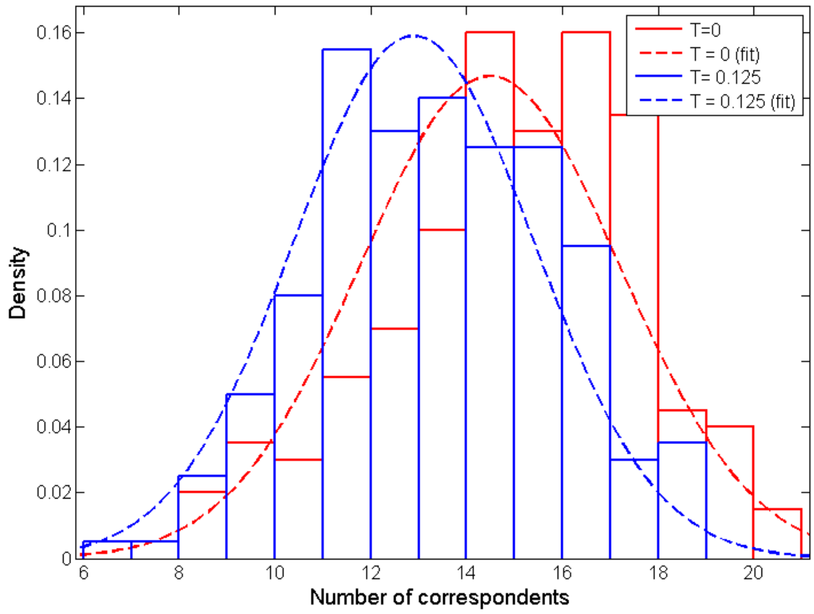

Figure 3: Correspondents - number

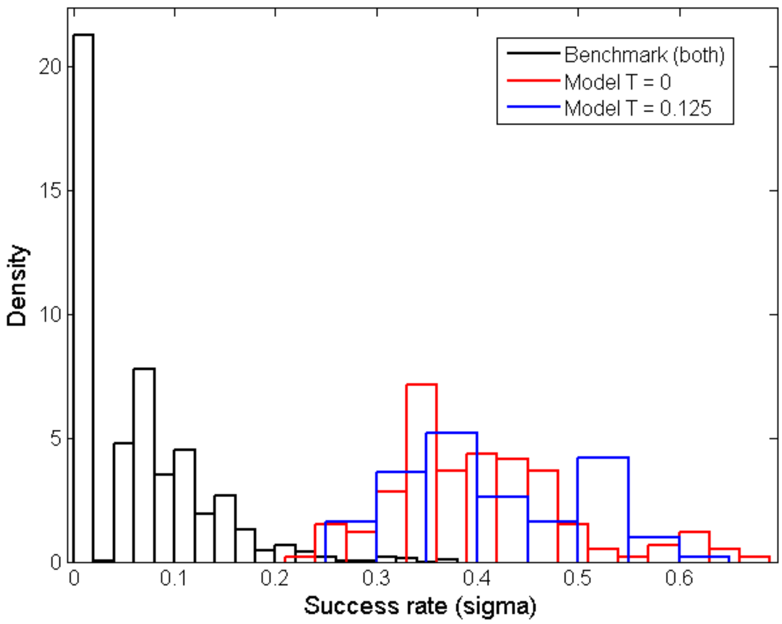

Figure 4: Correspondents - correctly identified (pdf)

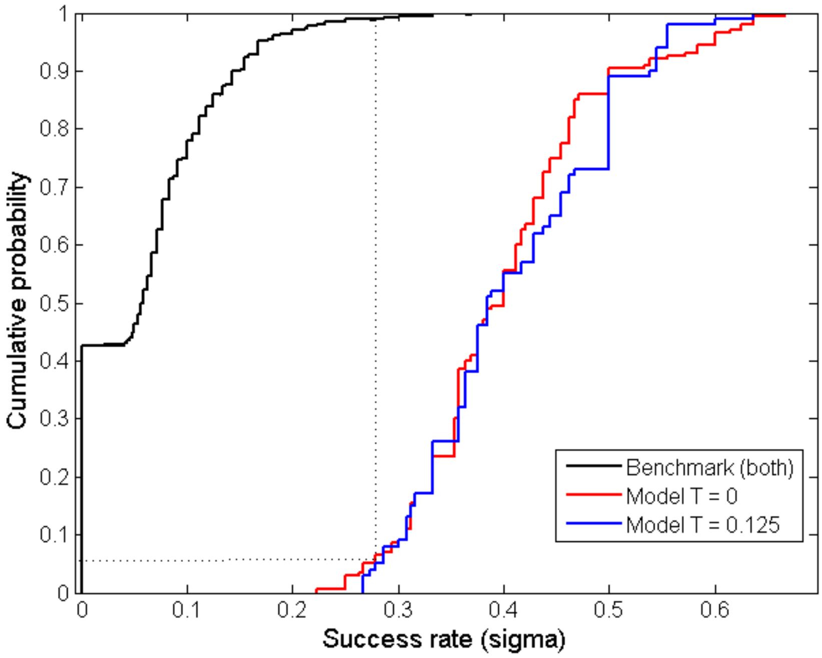

Figure 5: Correspondents - correctly identified (cdf)

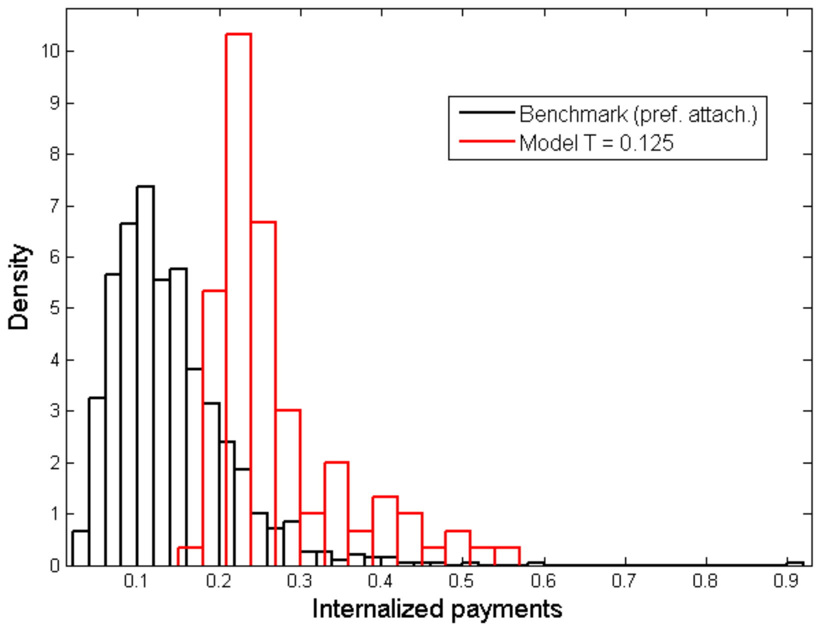

Figure 6: Internalized payments

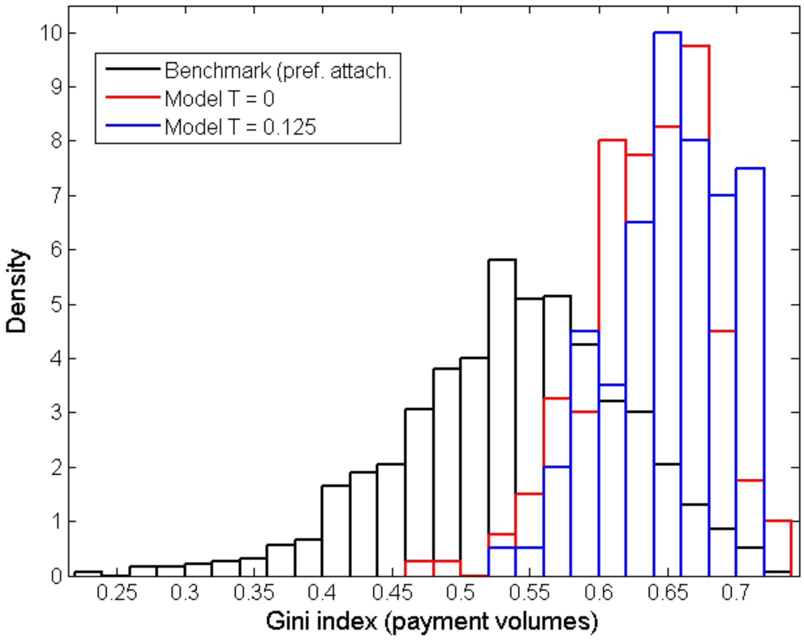

Figure 7: Concentration of payment volumes

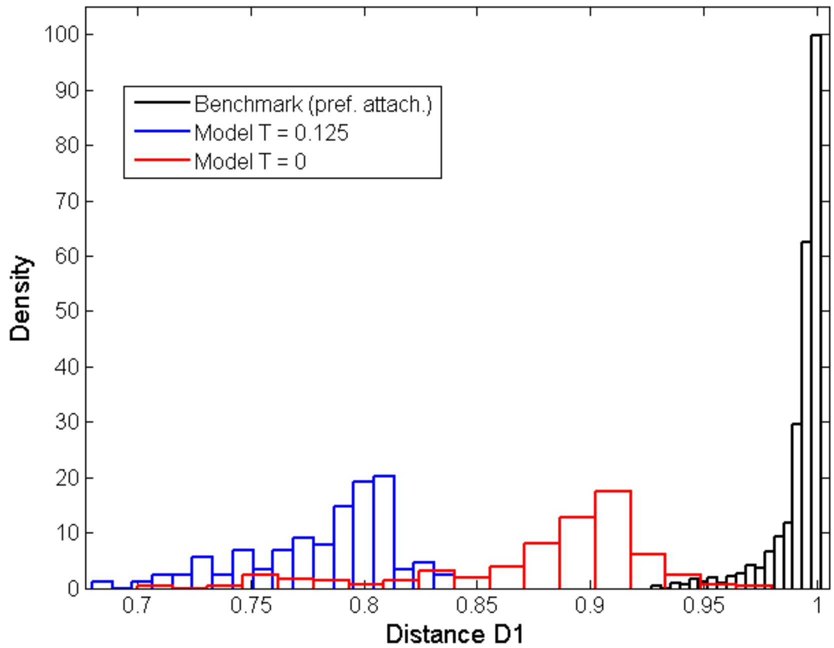

Figure 8: Link-change distance

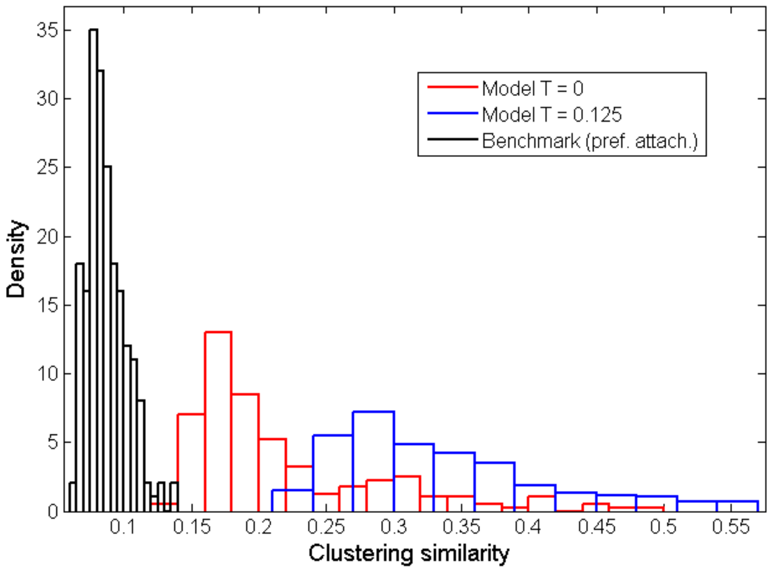

Figure 9: Clustering similarity

DEPFID WORKING PAPERS

DIPARTIMENTO DI POLITICA ECONOMICA, FINANZA E SVILUPPO
UNIVERSITÀ DI SIENA
PIAZZA S. FRANCESCO 7 I- 53100 SIENA
http://www.depfid.unisi.it/WorkingPapers/
ISSN 1972 - 361X