

An agent-based model of payment systems

Marco Galbiati[†] and Kimmo Soramäki[‡]

14 October 2007

Abstract

We lay out and simulate a multi-agent, multi-period model of an RTGS payment system. At the beginning of the day, banks choose how much costly liquidity to allocate to the settlement process. Then, they use it to execute an exogenous, random stream of payment orders. If a bank's liquidity stock is depleted, payments are queued until new liquidity arrives from other banks, imposing costs on the delaying bank. We study the equilibrium level of liquidity posted in the system, performing some comparative statics and obtaining: i) a liquidity demand curve which links liquidity to delay costs; ii) insights on the efficiency of alternative system configurations; iii) insights on the effects of operational incidents on the amount of liquidity present in the system.

[†] Bank of England. E-mail: 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.

[‡] Helsinki Univ. of Technology. E-mail: kimmo@soramaki.net. Contact information at: www.soramaki.net.

The authors thank M. Bech, W. Beyeler, M. Manning, S. Millard, F. Renault, M. Willison for important insights. Useful comments came from the participants of conferences and seminars: FS seminar at the Bank of England, Central Bank Policy Workshop (Basel), Computer Science Department seminar (University of Liverpool), CCFA Summerschool (University of Essex), 5th Simulator Seminar (Bank of Finland), Sixth International ISDG Workshop (Rabat, Morocco). The usual disclaimer applies.

1 Introduction

Virtually all economic activity is facilitated by transfers of claims towards public or private financial institutions. In turn these claim transfers generate payments between banks whenever they are not settled on the books of a same bank ("on us"). These payments are settled in interbank payment systems. In 2004, transfers in the EU interbank payment system, TARGET, totalled \$552 trillion on an annual basis. In the US Fedwire system the amount transacted was \$470 trillion, and in the UK CHAPS \$59 trillion - around 40 times the value of the respective gross domestic products (BIS 2006). The sheer size of the transfers, and their pivotal role for the functioning of financial markets and the implementation of monetary policy, make payment systems a central issue for policy makers and regulators.

At present, most interbank payment systems work on a real-time gross settlement (RTGS) modality. That is, settlement takes place as soon as a payment is submitted into the system (real time) and the corresponding full amount in central bank money (gross settlement) can be delivered. Because no netting takes place, RTGS modality imposes high liquidity demands on the banks, making RTGS vulnerable to *liquidity risk*, i.e. to the risk that liquidity-short banks are unable to send their own payments, creating delays and possibly gridlocking the whole system (see e.g. Bech and Soramäki 2003). It is fair to say that liquidity is *the main issue* in RTGS payment systems; as such it attracts the attention of central banks and stimulates a large amount of research. This paper aims at contributing to this knowledge, offering a model of liquidity demand and circulation in an RTGS system.

The amount and the distribution of liquidity in a payment system is the result of a complex interaction between the system's participants. During the day, each settlement bank receives a stream of payment instructions mandated by its clients¹ that is largely out of the bank's control and which, moreover, displays some randomness. To cover for the liquidity needs generated by these instructions, banks typically rely on two sources: a) reserve balances or credit acquired from the Central Bank and b) funds received from other settlement banks in the course of the day. The first source can be seen providing *external* (to the system) *liquidity* while the second is a source of *internal liquidity*. In normal conditions a bank can freely draw on external liquidity; this however has a cost, which gives incentives to economize on its use². Internal liquidity on the other hand carries no cost, but its arrival is out of the bank's control. Hence, reliance on internal liquidity exposes the bank to the risk of having to delay its own payment activity - something which also is costly. Thus, the problem of a bank is to make an optimal use of external liquidity, trying to forecast when and how much it will receive internal liquidity, and trading off external liquidity costs against (expected) delay costs.

¹Banks also make payments related to their own operations in the markets, but we leave these aside in this paper.

²Liquidity costs are opportunity costs (returns that the bank would obtain if it could employ this liquidity differently), of interest costs (costs from borrowing the liquidity itself).

Two main difficulties emerge when studying the behaviour of banks in a payment system. First, when modelled in sufficient detail, fluxes in RTGS systems follow complex dynamics, making the bank's liquidity management problem anything but trivial. Indeed, recent work by Beyeler *et al.* (2007) shows that, when the level of external liquidity is low, payments become coupled across the network and lose correlation with the arrival of payment orders; as a consequence, it is difficult to gauge the precise relationship between liquidity and delays, making it hard to determine the optimal usage of external funds. Second, the actions of each bank produce spillovers on the rest of the system, so no system participant can solve its optimal liquidity demand problem in isolation. In a word, strategic interactions are widespread, so banks interact in a fully fledged "game", jointly determining the performance of the system.

This paper studies this liquidity game, putting special effort into modelling liquidity fluxes. We thus build a payments model where external liquidity is continuously "recycled" among many banks, with delays and costs generated in a non-trivial way by a realistic settlement process. Such realism will inevitably force us to abandon the analytical way and to instead use simulations. In particular, we use numerical methods to calculate a crucial element of the game, the payoff function, or a relationship between i) a bank's own external liquidity, ii) the external liquidity of "the others", and iii) the resulting settlement delays and costs. Exploring the payoff function we find that, depending on the model specifications, banks engage different types of games, some of which turn out to be well known objects in game theory.

Being interested in the performance of the system, we are interested in the equilibria of the liquidity game, or the choices that banks may be seen to adopt in a consistent fashion. To do so we solve the model adopting a dynamic approach. That is, we assume that banks change their actions in time, using an adaptation process which prescribes new actions on the basis of past experience. We then simulate the resulting dynamics and we look at the limit behaviour. The limit, or equilibrium behaviour depends on the specific form of the adaptation rule, so we choose this learning process in such a way that, on the one hand, it embeds some rationality on the part of the banks; on the other, it leads to a meaningful equilibrium.³

Given its game-theoretic approach, this paper is related to recent work by Angelini (1998), Bech and Garratt (2003, 2006), Buckle and Campbell (2003) and Willison (2004). These papers model various "liquidity management games" with a few agents and a few periods (typically and respectively, two and three). While these models enhance our understanding of the incentives in payment systems, the actual payoff functions may be too simple to accurately describe costs in real payment systems. As we said, in RTGS systems liquidity can circulate many times and between many banks, generating dynamics that cannot be captured by these simple, analytically tractable, models.

Recently, a growing literature has used simulation techniques to investigate

³Slightly more technically, a convergence point of our dynamics will be a Nash equilibrium of the liquidity game.

efficiency and risk issues payment systems (see e.g. BoE (2004), the volumes edited by Leinonen (2005, 2007), and Devriese and Mitchell (2005)). Simulation studies have been widely used in comparing alternative central bank policies, or testing the impact of new system features before their implementation in payment systems. A common shortcoming of such studies has been, however, that participant behaviour is rarely endogenised in the models. The behaviour of banks has then either assumed to remain unchanged across alternative scenarios, or to change in a predetermined manner, leaving aside (or largely simplifying) the strategic aspects studied by the game-theoretic studies.

Recognizing the strengths, but also the limitations, of these two approaches, the present paper tries to build a bridge between them, aiming at joining what we see as convenient features of both of them.

The paper is organised as follows: Section 2 develops a formal description of the model, describes some properties of the cost function, and illustrates the tâtonnement process towards equilibrium. Section 3 presents the results of the experiments and section 4 concludes.

2 Description of the model

The model is a stylized representation of a day in RTGS; its agents are a number of banks, which engage in the following game.

2.1 Banks and external liquidity

At the beginning of the day, each of N banks (denoted by $i = 1 \dots N$) chooses its reserves, say $l_i(0)$, to be used in the course of the settlement day. We assume $l_i(0)$ is an integer between 0 and some large L . We call the vector $l = (l_1(0), l_2(0) \dots l_N(0))$ an action profile, with $l_i(0)$ the action of bank i . Once actions have been (simultaneously) chosen, the settlement day begins: banks start receiving payment orders, they execute them using available liquidity, obtaining an outcome determined as follows.

2.2 Payments and delays

The day is a sequence of discrete time ticks, indexed $t = 0 \dots T$. At any time t , with probability p bank i receives an instruction to pay 1 unit of liquidity to any other bank j . The number of payment orders received by bank i up to time t is called $z_i(t)$. Instead, $x_i(t)$ denotes the payment orders *executed* by i up to t . For bank i , the backlog of outstanding orders at t (bank i 's queue) is therefore

$$q_i(t) = z_i(t) - x_i(t),$$

We set $z_i(0) = x_i(0) = 0$. Payments orders are executed using available liquidity; bank i 's available liquidity at time t is defined as:

$$l_i(t) = l_i(0) - x_i(t) + y_i(t)$$

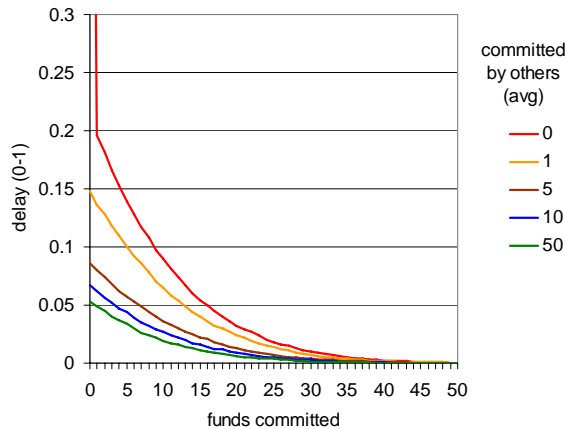


Figure 1: Delays as a function of total liquidity

where $y_i(t)$ is the amount of payments that i has received from other banks up to time t (so $\sum_i y_i(\cdot) = \sum_i x_i(\cdot)$). For simplicity, we assume that every i adopts the following payment rule:⁴

$$\begin{aligned} &\text{if } l_i(t) > 0, \text{ execute new and queued payments as FIFO;} \\ &\text{if } l_i(t) = 0, \text{ queue new payment orders.} \end{aligned} \quad (1)$$

Bank i 's incoming payments $y_i(t)$ are just "other banks'" outgoing payments, so the description of the settlement process is completed.

As mentioned in the Introduction, even this simple model generates extremely complex dynamics of liquidity $l_i(t)$ and queues $q_i(t)$.⁵ However, the model can be simulated numerically: a given action profile $l = (l_1(0), \dots, l_N(0))$ pins down the initial conditions of the system; from there, the exogenous arrival of payment orders mechanically generates liquidity fluxes, queues and delays. For example, Figure 1 shows the (average) amount of delays obtained for different levels of total liquidity in the system, when $l_i(0)$ is constant across i .

2.3 Costs

The settlement algorithm described above takes as an input the action profile $l = (l_1(0), \dots, l_N(0))$, and produces as an outcome a random amount of payments,

⁴Such rule is optimal for the cost specification given in the next section: banks need to pay upfront for liquidity, so they have no incentive to delay payments if liquidity is available. Under other cost specifications (e.g. priced credit or heterogeneous payment delay costs) this would, however, not be the case.

⁵Only when $l_i(0)$ is "very large", queues do not form, and $\Delta x = \Delta z$ so payments essentially follow a Poisson process which mirrors the arrival of payment orders.

delays and queues. In turn, this outcome brings about payoffs (or costs) to the players. We suppose that a payment order received at t and executed at t' carries a penalty equal to

$$c(t', t) = \kappa(t' - t), \quad \kappa > 0 \quad (2)$$

At the end of the day, such penalties are summed over all received payment orders, to give a bank's delay cost. Similarly, we suppose that acquiring initial liquidity $l_i(0)$ imposes a liquidity cost equal to:

$$C(l_i(0)) = \lambda l_i(0), \quad \lambda > 0 \quad (3)$$

A bank's total payoff is then the sum of delay and liquidity costs.

Because the arrival of payment orders is random, total payoffs are a random function of the action profile l . As anticipated in the Introduction, analytical derivation of the probability distribution of costs is exceedingly complex; hence, we simulate the settlement process many times for every action profile, to get a numerical estimate of the relationship between actions and expected costs.⁶ The resulting payoff function is plotted in Figures 2 and 3 (with delays more costly in the latter).

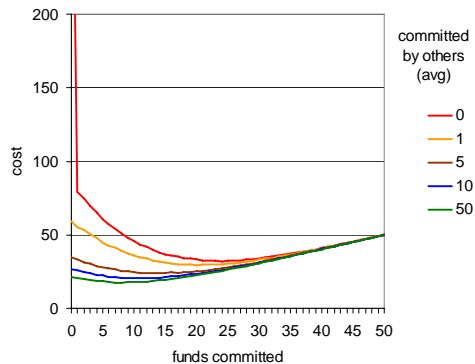


Figure 2: Costs as a function of own initial funds

An essential finding is that bank i 's cost is essentially a function of its own action and of the *sum* of others' actions; that is, total expected costs can be written $\pi_i(l) \approx f(l_i(0), \overline{l_{-i}(0)})$, with $\overline{l_{-i}(0)} = \sum_{j \neq i} l_j(0)$. This greatly simplifies the analysis because (in certain respects) it leaves us with a game with

⁶We suppose banks are risk-neutral: they care about expected payoffs.

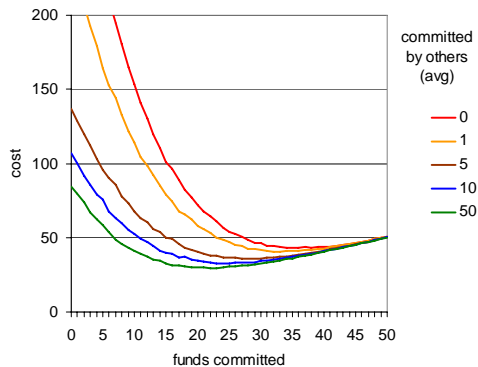


Figure 3: Costs as a function of own initial funds

only two players: bank i playing against "the rest of the system".⁷ We find this result by comparing two sets of simulations: one in which we initialize the settlement day varying the total amount of liquidity, spreading it uniformly across banks (i.e. we simulate a settlement day for different values of $\sum l_i(0)$, imposing $l_i(0) = \frac{1}{N} \sum l_i(0) \forall i$). In the other one, we look again at different levels of total initial liquidity, but we distribute it randomly across banks, so $l_i(0)$ varies across is . Comparing the total costs in the two settings, we found that the differences are small - around 2% or less. We suspect this stems on two facts: i) the assumption of a complete symmetric network (every bank exchange payments to any other with similar intensity), and ii) the relatively large number of payments to allow liquidity to be redistributed and flush out the initial conditions. Both assumptions are very realistic in several instances, for example in the UK CHAPS system.

2.4 Equilibrium

To find the equilibrium of the liquidity game, we use the so called *fictitious play* tâtonnement process. Largely studied in evolutionary game theory, fictitious play is a specification of how players change their actions in time, learning from experience. A precise description of this process is in Appendix; the reason why we adopt this particular dynamics is twofold. First, despite its simplicity the fictitious play rule is in a sense rational, corresponding to Bayesian updating

⁷Games with this property are known as "aggregation games", and have been studied in a game theory (see eg. Mezzetti *et al.* (2006) and the references therein). Among others, these games have the convenient feature that a number of adjustment dynamics applied to them are "well behaved". This will be useful searching the game's equilibrium - see next Section.

of beliefs about others' actions⁸; hence, the resulting story of banks repeatedly interacting is *per se* not too unrealistic. Second, fictitious play may be seen as a tool to compute the equilibria. When fictitious play converges to a stable action profile, this is a Nash equilibrium of the underlying game.⁹ So, depending on the point of view, fictitious play can be both seen as a computational device, or as a "story" with an appealing economic meaning.

A slightly technical but essential question is whether the game has a unique equilibrium and, if not, which equilibrium we uncover with fictitious play. The Appendix discusses this in more detail. The bottom line is that, although our model does have different equilibria (depending on the initial conditions of fictitious play we pick one or the other), all equilibria are characterized by a same total level of liquidity $\sum_i l_i(0)$, which can therefore be rightly called *the* equilibrium liquidity. This allows us to perform comparative statics changing parameters of the cost function and other elements of the model.

3 Results

3.1 Liquidity demand and efficiency of the equilibrium

We start with a base case scenario with 15 banks and 200 payments per bank. In all of the experiments banks interact in a complete topology, i.e. send and receive payments from each other bank in the system. Interested first in the amount of external liquidity as a function of unit delay costs, we keep λ (the price of external liquidity) normalized to 1.

As expected, we find that the amount of external liquidity acquired by the banks is very low when delays are relatively inexpensive (Figure 4, pg. 9). When k grows, so does liquidity at the system-level, which is seen to follow an exponential pattern up to a certain point. However, decreasing returns from liquidity eventually prevail¹⁰, and so liquidity demand eventually flattens out (liquidity is abundant when unit delay costs at very high).

Figure 5 reports the delays experienced in the system at different levels of (equilibrium) liquidity holdings, showing that systemic delays decrease exponentially as liquidity is increased, and vice versa. We are here experimentally recovering the phenomena often referred to as "deadweight losses" (Angelini 1998) or "gridlocks" (Bech and Soramäki 2002): a bank that reduces its liquidity holdings might have to delay its outgoing payments. As a consequence, the receivers of the delayed payments may in turn need to delay their own payments, causing further downstream delays and so on. Hence, a decrease of a unit of liquidity may cause multiple units to be delayed. Such a chain of delays - and hence this multiplicative effect - is more likely and longer, the lower the liquidity

⁸ See e.g. Fudenberg and Levine (1998) pg. 31 for details.

⁹ It is also well known that fictitious play may fail to converge. However this is not the case here, as shown by the simulations. At the root of this, the fact that we have an aggregation game with strategic complements (see Kukushkin (2006)).

¹⁰ As liquidity grows, delays become increasingly rare.

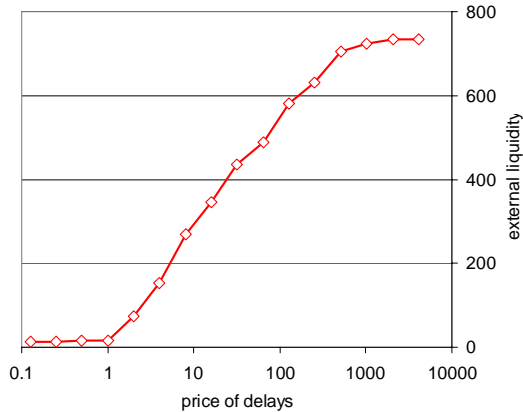


Figure 4: Equilibrium external liquidity as a function of delay costs

possessed by the banks. Thus the total effect of liquidity reduction acts in a compounded (exponential) fashion.

An important question is whether the equilibrium of the liquidity game is efficient. That is, whether the unregulated interaction of banks can be improved upon by means of proactive regulation. To answer this question, we could compare the equilibrium outcome of our game, to what would result if banks were *jointly* minimizing the *total* costs of the system. As a first, computationally simpler step, we search for an optimal liquidity level under the constraint $l_i(0) = l_j(0) \forall i, j$ (all banks choose the same amount of funds). We find that the equilibrium outcome roughly coincides with the collective-cost-minimizing choice for extreme values of k (delay costs), as shown in Figure 6, pg.10. However, for intermediate unit delay costs, the outcome reached by competitive banks is dominated by the coordination outcome, where more liquidity is provided as a whole. In these cases, competitive banks are seen to free-ride on the liquidity injected by other banks, leading to insufficient provision of external liquidity¹¹. At these delay cost levels, the strategic situation can be characterized by the classic hawk-dove game (further details in Appendix).

3.2 Relative efficiency of different size networks

Is a system with more participant preferable to a smaller one? This question can be considered from different points of view: from a risk and stability per-

¹¹We are unable to find *the* optimal level of liquidity, which could in principle be even lower than the Nash-equilibrium level. However, it is still true that banks would be collectively better off if they committed more liquidity than in the Nash equilibrium.

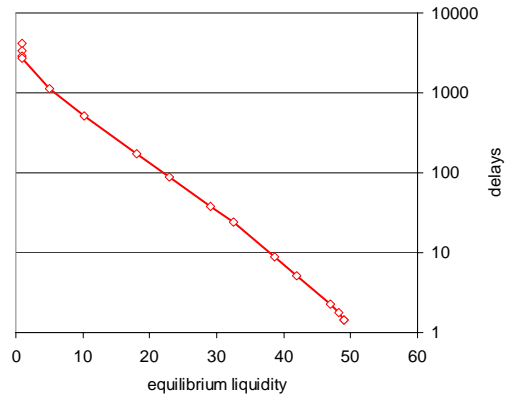


Figure 5: Delays as a function of equilibrium liquidity - horizontal axis: $\sum l_i(0)$; vertical axis: average total delays.

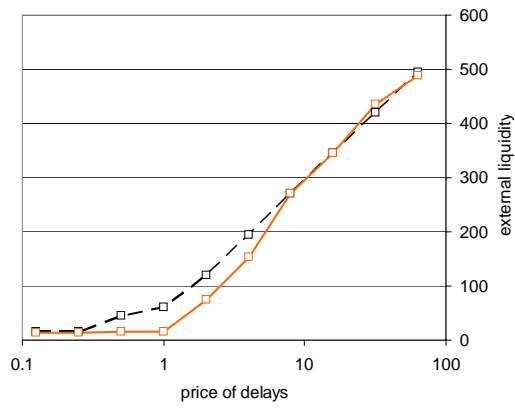


Figure 6: Cost-minimizing common action (black) *vs* Nash-equilibrium outcome (red)

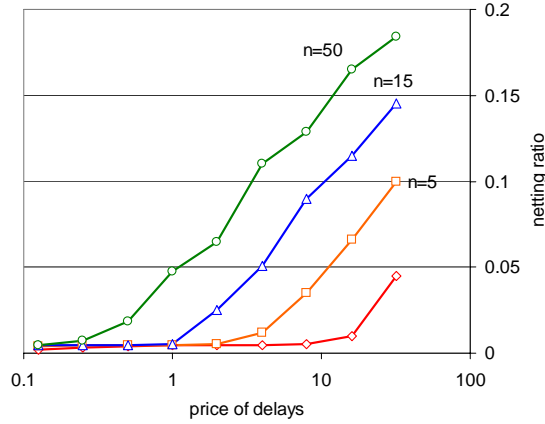


Figure 7: Equilibrium external liquidity with alternative system sizes and fixed turnover per bank

spective,¹² or from a cost-efficiency perspective. Here, we concentrate on the second aspect. We then run experiments varying the number of participants looking at the equilibrium levels of liquidity, delays and costs. In a first set of experiments, we keep the number of payments per bank constant. In a second scenario, we keep constant the system-wide volume, spreading it over different numbers of banks.

3.2.1 Size effects I - constant individual bank's payments

As expected, we observe that the comparison across networks of different sizes is not unambiguous; instead, differences across sizes in liquidity demand, depend on the level of unit delay cost. As Figure 7 shows, for low delay costs the netting ratio¹³ is unaffected by the network size. Instead, at higher unit delay costs, differences are amplified and small systems appear to be increasingly more efficient than larger ones.

The reason behind this result could be the following. Consider the service that a unit of external liquidity provides to a bank, in terms of reducing delays. This is positively related to the velocity at which the unit comes back as *internal* liquidity, after it has been circulated in the system. The higher this velocity, the smaller the amount of external liquidity needed to settle payments. Now, this velocity is smaller in larger systems: with many banks, when A sends a unit of liquidity to B, B is less likely to send it back to A, being instead more likely

¹²For example, fewer participants could imply higher risk, as the failure of a bank may mean that a larger share of payments is disrupted.

¹³The average liquidity required for each payment, or the ratio (tot external liquidity) / (tot payments).

to send it to some third bank C. Liquidity sent out by a bank is thus dispersed in large systems, taking on average a longer time to return to the originator. Liquidity has a longer return time in larger systems, which therefore "recycle" liquidity less efficiently.

3.2.2 Size effects II - constant total volume

In this second experiment, we keep constant the system total volume, distributing it over a varying number of banks. The results, illustrated in Figure 8 pg. 13, show a pattern similar to the previous case.

The hypothesis that more concentrated systems may be more liquidity-efficient, has already been formulated in the literature. One of the forces behind this phenomenon is the so-called "liquidity pooling" effect (see e.g. Jackson *et al.* 2007), which we see here at work.¹⁴ A rough description of liquidity pooling works in our context is the following. First, note that if the number of participants is increased by a factor x , the volatility of the balance of each bank is multiplied by a factor $1/x' > 1/x$ - we show this in a moment. Second, suppose that i) the optimal a_i is proportional to the volatility of a bank's balance δ (i.e. $a_i = z\delta$)¹⁵ and that ii) banks post all the same amount of liquidity (i.e. $a_i = a_j$). It then simply follows that the total amount of liquidity increases with the system's size: $(Nx) z \frac{\delta}{x'} > Nz\delta$ (here Nx is the number of banks in the larger system, and $\frac{\delta}{x'}$ the corresponding volatility of balances). The key point is that, if the number of participants is increased (keeping turnover constant), the volatility of banks balances rises more than proportionally. To see why this is the case, consider the simplified but illustrative situation where liquidity is abundant, so there are no delays. In this case, a bank's net position is the sum of a series of random perturbations (incoming and outgoing payments), equally likely to affect it positively and negatively. In other words, a bank's net position is a random walk, whose value after t perturbations averages zero, with a standard deviation \sqrt{t} . By increasing the system size by a factor x , the orders are distributed over more banks, so the average number of perturbations for any given time interval is multiplied by $1/x < 1$. Accordingly, the standard deviation of the balances at the end of any time interval is multiplied by $1/x' = \sqrt{1/x} > 1/x$.

3.3 Liquidity choices in the presence of operational incidents

Outages by banks in the payment system are rare in actual payment systems, but do take place occasionally. In a typical scenario a participating bank experiences problems connecting to the system due to temporal unavailability of IT

¹⁴Another reason for liquidity savings is the "internalization effect", which emerges when payments are settled "on us". This is absent here, as we keep constant the number of payments settled *across the system*.

¹⁵This is exactly the case if a bank chooses a_i as to cover z "standard deviations" from the average balance.

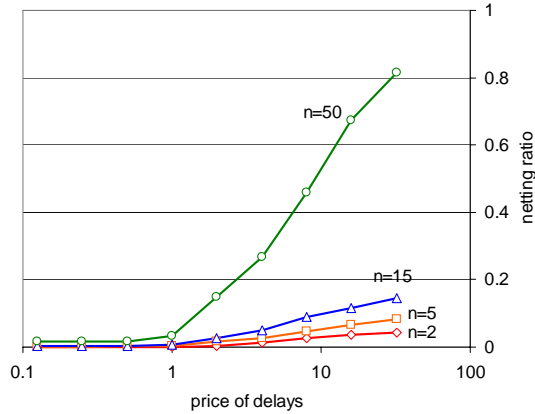


Figure 8: Equilibrium external liquidity with alternative system sizes and fixed total turnover.

systems or telecommunication facilities. In such situations it is frequent that a disconnected bank may still receive payments, but it cannot submit instructions to pay from its account. Unless the other banks stop paying to the troubled bank, this latter quickly becomes a liquidity sink, subtracting external liquidity from the system and endangering the system's viability.¹⁶

In a last set of simulations we ask the question of how much liquidity banks would wish to commit in such a situation, i.e. what is the impact of an operational outage on the demand for intraday credit. The banks are assumed to be unaware of the possible incident, and unable to discriminate among their counterparts, so the intraday liquidity management rule under 1), pg.5, is still adopted. Under these assumptions, we simulated a scenario where a randomly selected bank can receive, but cannot send payments for the first half of the settlement day¹⁷.

The findings of this experiment are summarized in Figure 9. The effect of an operation incident on the funding decisions by the banks is found to be rather limited, and ambiguous depending on the delay costs. At very low delay costs, banks prefer to hold about the same amount of liquidity as without the incident, implicitly choosing to experience the delays caused by the reduced liquidity. At very high price of delays bank choose to use less liquidity. We still don't have a clear explanation for this non-monotonic pattern.

¹⁶see e.g. analysis on the impact of the 9/11 terrorist attacks in McAndrews and Potter (2002), Lacker (2004), and Soramäki et al (2007)

¹⁷This generates a liquidity sink whose size equals the number of payment orders received by the distressed bank, and not yet executed until the second half of the day. This ranges between 100% and 14% of the system's external liquidity, depending on the level of unit delay costs.

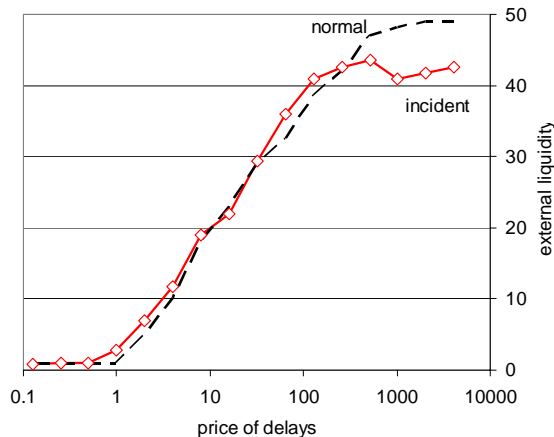


Figure 9: Equilibrium external liquidity with operational incident

4 Conclusions

In this paper we build and simulate a model of an RTGS system, paying special attention to the complex liquidity fluxes exchanged by the participating banks. The simulations show that a complete, symmetric RTGS system can be described as an "aggregation game", whose convenient features allow us to compute the equilibrium behaviour of the settlement agents.

We can thus retrieve a liquidity demand function which relates the liquidity present in the system, to the costs (liquidity vs delay costs) faced by banks in their payment activity. We then consider the question of whether the equilibrium reached by independent, competitive banks is efficient. On the one hand we find that, for a wide array of cost parameters, total settlement costs could be *reduced* if banks were to commit *more* liquidity than they do in equilibrium. This might constitute a rationale for the imposition of throughput guidelines, or of other measures which encourage the system members to inject liquidity in the system. From a different perspective, we find that systems with fewer participants are more liquidity-efficient than larger system, observing the emergence of "liquidity pooling" effects hypothesized by previous studies (Jackson *et al.* (2007)). Finally, we evaluate the incidence of operational incidents on the amount of liquidity that the banks are willing to inject in the system. We found that the existence of temporary liquidity sinks has a (surprisingly) small impact on the banks' liquidity choices.

Interesting further research may look at different network structures, at non-homogeneous bans (which could be facing different costs and different randomness in payment orders), at more elaborated intraday liquidity management rules.

5 Appendix

5.1 Fictitious play

Suppose that the daily game is repeated in time, thus giving rise to a sequence of "settlement days" running from $t = 0$ to (potentially) infinity. The action profile chosen on day t is denoted by $l^t = \{l_1^t, l_2^t, \dots, l_N^t\}$ which, as far as costs are concerned, can be summarized into $l^t = (l_i^t, \overline{l_{-i}^t})$.¹⁸

Fictitious play assumes that, over the sequence of days, every i forms a belief of what others will play next, choosing l_i^t as a best reply to such belief:

- i 's belief at time t is a vector of $\tilde{p}_i^t(\cdot) = (\tilde{p}_i^t(1), \tilde{p}_i^t(2) \dots)$, where $\tilde{p}_i^t(x)$ is the probability that i attaches to $\overline{l_{-i}^t} = x$.
- a belief is modified according to the following rule:

$$\tilde{p}_i^t(k) = \frac{1 + \sum_{s=1 \dots t-1} I_k(s)}{t + \Lambda}$$

where $\Lambda = NL$ (N being the number of banks, L the maximum liquidity each can post), and $I_k(s)$ is defined to be 1 if $\overline{l_{-i}^s} = k$, and zero otherwise.¹⁹

- i chooses $l_i^t = \arg \max_l \sum_{x=1}^L f_i(l, x) \tilde{p}_i^t(x)$ - recall that $f_i(l, x)$ is the cost incurred by i by playing l if others play x .

5.2 Equilibria of the liquidity game

The equilibria that we find are mostly mixed where banks switch between two or more actions depending on the evolution of their beliefs. At the lowest delay price level banks mix between no liquidity and one unit of liquidity with probabilities 8.6% and 91.4%, respectively. As banks become sufficiently confident that other banks commit one unit of liquidity each, the best reply is to commit no liquidity. As the probability of committing no liquidity is thereby increased, banks switch back to committing one unit of liquidity. The game can be characterized as a classical hawk-dove game. If no-one commits any liquidity all will experience very high delays as no payments can be settled. If everyone commits one unit of liquidity, payment settlement can take place. From an individual banks perspective, however, a better outcome would be not to commit any liquidity while others do. As the cost for delays is increased, the probability of banks committing no liquidity is reduced gradually until at delay price of one, a unique equilibrium of each bank committing one unit of liquidity exists. At higher cost levels bank either find a unique equilibrium or a mixed equilibrium where they mix between a narrow range of different liquidity levels.

¹⁸Here l_i^t denotes the action $l_i(0)$ chosen at time zero in day t . We are not interested in the intraday timing now, but rather in sequence of days, so we slightly change notation.

¹⁹On the first day ($t = 0$), all banks believe that each $\overline{l_{-i}^0}$ is equally likely ($\tilde{p}_i^0(k) = 1/\Lambda$). To start with different beliefs, one can change the parameter Λ .

5.3 Uniqueness of equilibrium liquidity level

We now provide an argument why all equilibria feature the same level of aggregate liquidity.

Given action profile l , call $l_i^*(l)$ bank i 's best reply to others' actions, l_{-i} . Given two action profiles l and l' , define $\Delta l_i^* = l_i^*(l') - l_i^*(l)$ (and similarly $\Delta \overline{l}_{-i} = \overline{l}'_{-i} - \overline{l}_{-i} = \sum_{j \neq i} l'_j - \sum_{j \neq i} l_j$). We now proceed in three steps.

Step 1) Delays are a weakly concave function of own liquidity. This emerges very clearly from the simulations; it is also intuitive considering the sequential nature of the settlement process (a bank's units of liquidity are called into play sequentially once previous units are released).

Step 2) For any two action profiles l and l' we have

$$\Delta l_i^* \leq -\Delta \overline{l}_{-i}$$

Indeed, suppose the "others" post x less. Then, by posting x more a bank more than compensates this liquidity reduction. The reason being that a bank values others' liquidity only when it receives it, so buying it in the first place (as l_i) is most beneficial. Hence, if at action profile $(l_i^*(l), l_{-i})$ bank i suffers delays D , then at $(l_i^*(l) + \varepsilon - \delta, l_{-i} - \varepsilon)$ with $\delta < \varepsilon$ the delays are again D . But i minimizes costs at $(l_i^*(l), l_{-i})$, and so given decreasing returns in liquidity (Step 1) bank i 's best reply to $l_{-i} - \varepsilon$ is (weakly) less than $l_i^*(l) + \varepsilon - \delta$ - recall that marginal liquidity costs are constant.

Step 3) Suppose there were two NEs (l^* and l'^*) with $\sum l_i^{*'} > \sum l_i^*$. This means $\Delta l^* = \sum \Delta l_i^* > 0$, which is impossible by Step 2: $\Delta l_i^* + \Delta \overline{l}_{-i}^* \leq 0$ for each i .

6 Bibliography

References

- [1] **Angelini, P (1998)**, "An Analysis of Competitive Externalities in Gross Settlement Systems", *Journal of Banking and Finance* n. 22, pages 1-18.
- [2] **Bank for International Settlements (2006)**, Statistics on payment and settlement systems in selected countries - Figures for 2005 - Preliminary version. *CPSS Publication* No. 75.
- [3] **Bech, M L and Soramäki, K (2002)**, "Liquidity, gridlocks and bank failures in large value payment systems". in *E-money and payment systems Review*, Central Banking Publications. London.
- [4] **Bech, M L and Garratt, R (2003)**, "The intraday liquidity management game", *Journal of Economic Theory*, vol. 109(2), pages 198-219.
- [5] **Bech, M L and Garratt, R (2006)**, "Illiquidity in the Interbank Payment System Following Wide-Scale Disruptions". *Federal Reserve Bank of New York Staff Report* No. 239.
- [6] **Beyeler, W, Bech, M, Glass, R. and Soramäki K (2006)**, "Congestion and Cascades in Payment Systems". *Federal Reserve Bank of New York Staff Report* No. 259.
- [7] **Buckle, S and Campbell, E (2003)**, "Settlement bank behaviour in real-time gross settlement payment system". *Bank of England Working paper* No. 22.
- [8] **Fudenberg, D and Levine D K (1998)**, "*The Theory of Learning in Games*". MIT Press., Cambridge, Massachusetts.
- [9] **Jackson, J and Manning M J, (2007)**, "Central Bank intraday collateral policy and implications for tiering in RTGS payment systems", DNB Working Paper No. 129, January 2007.
- [10] **James, K and Willison M. (2004)**, "Collateral Posting Decisions in CHAPS Sterling", *Bank of England Financial Stability Report*, December 2006.
- [11] **Kukushkin, N S (2004)**, "Best response dynamics in finite games with additive aggregation", *Games and Economic Behavior* (48), 94-110.
- [12] **Lacker, J M (2004)**, "Payment system disruptions and the federal reserve following September 11, 2001." *Journal of Monetary Economics*, Vol. 51, No. 5, pages 935-965.
- [13] **McAndrews, J and Potter S M (2002)**, "The Liquidity Effects of the Events of September 11, 2001", *FRBNY Economic Policy Review*, Vol. 8, No. 2, pages 59-79.

- [14] **Mezzetti, C, Dindoš M (2006)**, Better-reply dynamics and global convergence to Nash equilibrium in aggregative games, *Games and Economic Behavior* (54), February 2006, Pages 261-292.
- [15] **Soramäki, K, Bech, M L, Arnold, J, Glass R J and Beyeler, W E (2006)**, "The Topology of Interbank Payment Flows". *Physica A* Vol. 379, pages 317-333.
- [16] **Willison, M (2005)**, "Real-Time Gross Settlement and hybrid payments systems: a comparison", *Bank of England Working Paper* No. 252.