Banks' Intraday Liquidity Management during Operational Outages: Theory and Evidence from the UK Payment System

Ouarda Merrouche* Jochen Schanz

Draft, May 2008†

Abstract

We investigate how settlement banks in CHAPS, the United Kingdom's large-value payment system, react to outages experienced by counterparties. If banks do not sufficiently monitor their outgoing payments, operational shocks can impact the entire payment system: the stricken bank absorbs liquidity. We first build a game-theoretic model in which a bank's decision to make payments depends on whether another bank experiences operational problems, and on the time of day at which the outage occurs. We then investigate these reactions empirically using a non-parametric method. Our theory predicts that banks stop paying to a stricken bank early in the day, when they are uncertain about the payment instructions they might have to execute. When this uncertainty has been resolved (later in the day), healthy banks make payments even to stricken banks. Both predictions are supported by the data. We show that this behaviour effectively contains the disruption caused by the operational outage: payment flows between healthy banks remain virtually uninterrupted.

1 Introduction

Payment and settlement systems are vital to the smooth functioning of any advanced economy. They are used to settle trades in foreign exchange, equities, bonds and money market instruments. Consumers rely on them to make house purchases, receive salaries and benefits, and pay for goods

*Bank of England, Threadneedle Street, London, EC2R 8AH. Emails: ouarda.merrouche@bankofengland.co.uk, jochen.schanz@bankofengland.co.uk.

†Please do not quote without the authors' written consent.

and services. We investigate how settlement banks in CHAPS, the United Kingdom's large-value payment system, react to outages experienced by another CHAPS settlement bank.[1] In liquidity-hungry systems, there is a risk that settlement banks continue to make payments to a bank that is able to receive but unable to make payments. The bank experiencing operational problems thereby involuntarily absorbs liquidity: it becomes a 'liquidity sink'. This liquidity is not available any more to execute payments between other, healthy settlement banks. Thus, if banks do not sufficiently monitor their outgoing payments, operational risk at one bank is a source of risk for the entire payment system: that is, a source of systemic risk.

We first build a game-theoretic model in which a bank's decision to make payments depends on whether another bank experiences operational problems, and on the time of the day at which the problems arise. In the empirical part, we estimate these reactions to an operational outage using data from CHAPS. Our theory predicts that banks stop paying to a stricken bank early in the day, when they are still uncertain about their payment flows. When this uncertainty has been resolved, healthy banks make payments even to stricken banks. Both results are supported by our empirical evidence. We show that this behaviour is sufficient to contain the effects of the operational problem in the sense that payment values between healthy banks remain unaffected. In addition, there is some limited evidence that banks cut their outgoing payments to a stricken bank by more in the second half of 2007, when market uncertainty was high. We discuss the extent to which our model helps to understand this result.

We hope to contribute to the existing literature in two respects. First, to our knowledge, this is the first paper to analyse how banks react to operational outages changes during the day. Second, we apply a more rigorous econometric approach than previous studies to analyse the frequency with which payments are made, following Engle and Russel (1998). The method we employ should be well suited to the analysis of high-frequency, irregularly spaced payment-by-payment data. In particular, we do not have to aggregate data in arbitrary intervals. We can rely on non-parametric methods that provide a thorough picture of changes in payment flows before, during and after outages.

The paper is organized as follows. Section 2 provides a brief overview of related literature. Section 3 presents the theoretical model; section 4 the empirical results. Section 5 concludes. A description of CHAPS can be found in section 4.1.

[1] Settlement banks are direct members of the payment system and settle payments on behalf of their clients (consumers, corporates and banks without direct membership). In the following, we will use the term settlement bank and bank interchangeably.

2 Related literature

Game-theoretic models of behaviour in large-value payment systems (such as CHAPS) predict that the timing of payments in real-time gross settlement systems is the result of banks trading off delay costs and liquidity costs. The argument runs as follows. Intraday liquidity can be drawn from two sources: (1) from the central bank (the settlement agent in CHAPS) against collateral; (2) from incoming payments. In the first case, the cost of liquidity is the opportunity cost of having to hold (and transfer) securities eligible as collateral. In the second case, banks may not receive sufficient payments in time to execute their payment instructions promptly; delay, however, could be expensive when contractual obligations or market practice are violated. As banks seek to minimize the cost associated with sending payments, their choice determines the distribution of payments throughout the day.

The starting point of our theoretical model is Bech and Garratt (2003). In their model, high liquidity costs encourage banks to delay payments, awaiting the receipt of incoming payments to fund their outflows. We retain their assumption that there are two banks that pay each other but increase the number of periods in which settlement banks can make payments to each other to three (morning, afternoon, and evening) to be able to describe the incentive to delay payments in the morning and the afternoon. To be able to analyse why they react differently to shocks in the morning and the afternoon, we further extend their analysis: we allow operational shocks to occur in each period; that banks do not know all their payment instructions at the beginning of the day; and we distinguish two types of payment instructions, 'normal' and 'urgent' ones.

Angelini (1998) considers the behavior of banks with both liquidity and delay costs in a RTGS system. His model's equilibrium involves excessive delay of payments, as banks do not internalize the benefits to other banks from the receipt of funds. Mills and Nesmith (2006) and Kahn et al (2003) consider the effect of settlement risks[2] on timing decisions. They illustrate another rationale for delays: uncertainty about whether the other participants might either default or delay can prompt the participants to delay their payments to obtain a better forecast of the cost of funding their own outflows.

A few, so far mostly descriptive empirical papers analysed payments data in normal and stressed environments. McAndrews and Rajan (2000) document the payment and value timing distribution of the Fedwire Funds Service[3] using data aggregated within ten minutes intervals. Becher et al (2007) carry out a similar descriptive analysis on CHAPS Sterling. McAndrews and Potter (2002)

[2] Here, settlement risk refers to the risk that a payment instruction that a client sends to a bank is not executed.

[3] The United States' large value RTGS payment system.

estimate the average bank payments reaction function in Fedwire using a panel fixed effect estimator on minute-by-minute data. Armentier et al (2007) evaluate the relationship between liquidity costs (proxied by payments values and volumes) and the timing distribution of Fedwire Funds transfers using hourly data.

3 Model

The model covers payments behaviour on a single day. Two banks decide at the start of the day how much liquidity to borrow from the central bank. In the subsequent periods, they decide whether to delay the execution of their payment instruction(s). Whether delay is attractive depends on how much liquidity each bank has available, on its opponent's strategy, and on whether operational shocks have hit one or both banks. The following section formalises the setup. We then provide some intuition for the trade-offs that banks face. Section 3.3 guides the reader through our results. The proofs are discussed in the appendix.

3.1 Setup

Two banks $i = 1, 2$ interact in four periods $t = 1, 2, 3, 4$. In the first, they simultaneously decide on their collateral postings $C_i \in \{0, 1, 2\}$ at cost γC_i and receive the instruction to execute a normal payment of value 1 to their opponent. Banks incur the one-off fee γC_i independently of how long they need the liquidity.[4] Collateral posting decisions remain private information. Three periods in which payments can be made follow. In each period, each bank can be hit by an operational shock $s_i^t \in \{0, 1\}$, where $t \in \{M, A, E\}$ indexes the periods, with probability ε_i. If $s_i^t = 1$, the bank is unable to make payments in this period, but able to receive them.[5] Shocks are publicly observable[6] and independently distributed across periods and banks. If $s_i^t = 0$, bank i can execute all payment instructions it has received if it has sufficient liquidity at the beginning of this period. That is, there is no possibility to net payments within a period. As in Bech and Garrat (2003), this assumption is made to reflect the key characteristic of any real-time gross settlement systems, that is, that payments cannot be netted.

[4] Virtually without exception banks post all collateral for the entire day before CHAPS opens, suggesting that the cost of intraday credit in CHAPS is independent of its duration.

[5] Notice that the receipt of payments is always possible unless the central bank's payment system breaks down. We do not consider this type of operational outage. Instead, we look at bank-specific shocks.

[6] In CHAPS, settlement banks are required to report operational problems within 15 [CHECK] minutes of the start of the outage.

At the start of the day, each bank obtains one payment instruction of value 1. In the second period, the afternoon, each bank may obtain an additional instruction of value 1 with probability v_i. In contrast to the morning instruction, this one is urgent, and delay to the evening costs $d > \gamma$. In the third period, the evening, no further instructions arrive. In the evening, each bank can attempt raise additional liquidity at cost γ to settle any outstanding instructions (unless it is hit by an operational shock). This attempt, however, fails with probability 1/2. Banks incur a cost of $f_n > \gamma$ (for normal instructions) and $f_u > \gamma, d$ (for urgent instructions) for any outstanding payment instruction that is not executed at all.[7] Each bank is trying to minimise the total costs arising from posting collateral and delaying / failing to execute payment instructions. There is no penalty for ending the day with a negative balance, nor a benefit for ending it on a positive balance.[8]

3.2 Trade-offs and intuition for main results.

At the start of the day, each bank has to decide how much liquidity it borrows from the central bank to settle its payments. We endogenise this decision; however, to understand the main trade-offs, it is useful to assume that this decision has been made, and investigate the bank's payment behaviour in the morning, the afternoon, and the evening.

- Suppose first that the bank has posted two units of collateral in the morning. Then it has no reason to delay: the expense for the collateral has been incurred, and liquidity suffices to make both payments. The bank might as well use it to make the payments to avoid the risk that it will not be able to do so in a later period, given that its systems may be hit by an operational shock. Thus, the risk of operational shocks, together with the cost of technical default f_n and f_u, means that it is, in expectation, costly to delay even normal payment instructions, and adds to the delay cost of the urgent payment instruction (d).

- Now suppose that the bank decided to post one unit of collateral. The incentive to execute payment instructions quickly remains. But if it executes the normal payment instruction in the morning, it may not have any liquidity left to execute the urgent payment instruction immediately in the afternoon. Hence, if the arrival of the urgent payment instruction in the

[7] By convention, f_u includes d: That is, if an urgent payment instruction is not executed, the cost is f_u, not $f_u + d$.

[8] This essentially implies that costs of (overnight) borrowing are zero. While this may appear inconsistent with the assumption of positive costs of borrowing intra-day, it simplifies notation. Our results go through as long as the cost of failing to make a payment is higher than the cost of borrowing overnight; and the penalty for not repaying the central bank at the end of the day is higher than the benefit of lending money out in the interbank market. Both assumptions are easily fulfilled in practice.

afternoon is sufficiently likely (v_i large), or its cost of delay high (d large) relative to the risk of operational failure ε_i, then the bank will prefer to save the liquidity for the afternoon if it does not anticipate to receive a payment from its opponent in the morning. If, in contrast, it expects to receive such a payment in the morning, it can use this liquidity in the afternoon to execute the urgent payment instruction. In this case, there is no benefit from delaying the normal payment.

Thus, optimal payments behaviour when the bank posts one unit of liquidity depends on its opponent's behaviour. An interesting case is when the opponent has not posted any collateral. Of course, he is then unable to pay in the morning, and relies on incoming liquidity to finance its payments. But this means that there is an additional benefit of paying early for the bank that posted one unit of collateral: if it makes the payment in the morning, the opponent can pay back in the afternoon, and the bank that posted one unit can re-use this liquidity in the evening. Of most interest is, however, the case in which both banks post one unit of collateral, and in which the delay cost d exceeds a threshold $d_{L,i}$. Then there may be two equilibria: one in which no bank pays in the morning, and one in which both pay unless one of them is hit by a shock. In either case, both banks have sufficient liquidity to execute an urgent payment instruction in the afternoon. Payoffs are lower in the first equilibrium because any delay increases the risk that payments may not be executed at all because of operational shocks.

- If the bank decides to post no collateral, it fully relies on incoming liquidity to make its payments, and / or on a successful attempt to raise liquidity in the evening. Clearly, this is only optimal when costs of liquidity are very high, operational shocks unlikely, the delay of urgent instructions inexpensive, and the costs of failure to execute payments low.

It is interesting to note that if the cost of delaying the urgent payment is high ($d > d_{L,i}$), then the decision between posting one or two units of collateral does not depend on d. This is because independently of opponent play, the bank will always prefer to ensure that it has sufficient collateral available to execute the urgent instruction immediately. For sufficiently high costs of collateral γ, the bank will only post one unit. In contrast, the decision between posting zero or one, and between posting zero or two units of collateral depends on d. For sufficiently high delay costs d, posting one or two units is preferred. Taken together, these features imply that for sufficiently high delay costs, and high costs of collateral, both banks post one unit of collateral in all equilibria.

The following section presents the setup and the main result more formally. Readers less interested in the game-theoretic modelling are invited to jump straight to the empirical results in section 4.

3.3 Equilibrium

The game is a finite two-player game. We are only looking for pure-strategy equilibria. The solution is via backwards induction. This section first considers equilibrium play starting from the afternoon period, and then moves backwards to the morning period, and the collateral posting decision. Let $p^t_{n,i} \in \{0,1\}$ be the number of normal payment instructions player i executes in period t, and correspondingly $p^t_{u,i} \in \{0,1\}$ for the urgent payment instructions. $p^t_i = p^t_{n,i} + p^t_{u,i}$ is the total number of payment instructions player i executes in t. Let l^t_i denote the available liquidity at the beginning of period t, that is, before any period-t payments are made or received. $\upsilon_i \in \{0,1\}$ denotes whether player i receives an urgent payment instruction. The intra-period timing is as follows for both players $i \in \{1,2\}$: i receives a payment instruction (not in period E); i learns whether he and / or his opponent is hit by an operational shock lasting for the entire period; (in period E only: i decides whether to attempt to raise additional liquidity); if i is healthy, he decides how many payment instructions to submit, subject to having sufficient liquidity available; i's cash account is debited with outgoing payments; i's cash account is credited with incoming payments.

Depending on the parameter values, different equilibria exist. Indeed, for a given set of parameter values, there may be multiple equilibria. We focus here on a specific set of parameter values:

1. The opportunity costs a settlement bank incurs when investing in (low-yielding) collateral and submitting it to the central bank are sufficiently high to discourage the settlement bank to post enough liquidity that would make it independent of any incoming payments with certainty. Formally γ must exceed for both players i a threshold $\gamma_{L,i}(C_j)$ whose value depends on the collateral C_j that the opponent j posts in the first period of the game. $\gamma_{L,i}(C_j)$ is defined in lemma 3 in the appendix.

2. The delay costs of the urgent transaction are assumed to be sufficiently high to discourage the bank from not posting any liquidity at all. Formally d must exceed a threshold $d_{L,i}(C_j)$ for both players i whose value depends again on C_j. $d_{L,i}(C_j)$ is defined in definition 1 at the start of the appendix.

For these ranges of γ and d, there exist two equilibria, and players find it optimal to post exactly one unit of collateral at the beginning of the day in both of them. Equilibrium behaviour in the afternoon and the evening is identical in both equilibria: In the afternoon, all available liquidity is used to execute outstanding payment instructions unless the player is hit by an operational shock. Normal instructions are only executed after urgent instructions have been executed. In the evening, all remaining instructions are executed subject to available liquidity unless the player is hit by an operational shock. If liquidity is insufficient, and $s_i^E = 0$, an attempt is made to raise additional liquidity.

But equilibria differ in their payments behaviour in the morning. In the first equilibrium, E1, neither bank makes a payment in the morning. In the second equilibrium, E2, banks pay each other in the morning unless they, or their opponent, suffer from an operational problem. Behaviour in E2 ensures that whether or not a bank or its opponent experiences an operational outage, it has sufficient liquidity available at the start of the second period to execute the urgent payment instruction immediately. Proposition 1 makes formally states the equilibrium.

Proposition 1 *If, for both players i, $\gamma > \gamma_{L,i}(1)$ and $d > d_{L,i}(1)$, then there exist two symmetric equilibria. In both these equilibria, $C_1 = C_2 = 1$, and payments in the afternoon and the evening are given by*

$$p_{u,i}^A = \left(1 - s_i^A\right) \min\left\{v_i, l_i^A\right\}$$

$$p_{n,i}^A = \left(1 - s_i^A\right) \min\left\{1 - p_{n,i}^M, l_i^A - p_{u,i}^A\right\}$$

$$p_{u,i}^E = \begin{cases} \left(1 - s_i^E\right)\left(v_i - p_{u,i}^A\right) & \begin{cases} \text{if } l_i^E \geq v_i - p_{u,i}^A, \\ \text{or } l_i^E < v_i - p_{u,i}^A \text{ and } i \text{ raised additional liquidity} \end{cases} \\ 0 & \text{if } l_i^E < v_i - p_{u,i}^A \text{ and } i \text{ could not raise additional liquidity} \end{cases}$$

$$p_{n,i}^E = \begin{cases} \left(1 - s_i^E\right)\left(1 - p_{n,i}^M - p_{n,i}^A\right) & \begin{cases} \text{if } l_i^E - p_{u,i}^E \geq 1 - p_{n,i}^M - p_{n,i}^A, \\ \text{or } l_i^E - p_{u,i}^E < 1 - p_{n,i}^M - p_{n,i}^A \text{ and } i \text{ raised additional liquidity} \end{cases} \\ 0 & \text{if } l_i^E - p_{u,i}^E < 1 - p_{n,i}^M - p_{n,i}^A \text{ and } i \text{ could not raise additional liquidity} \end{cases}$$

- *In equilibrium E1, $p_1^M = p_2^M = 0$.*
- *In equilibrium E2,*

$$p_1^M = \left(1 - s_i^M\right) \cdot \begin{cases} 1 & \text{if } p_j^M = 1 \\ 0 & \text{if } p_j^M = 0 \end{cases}$$

If, for all C_j, $\gamma > \gamma_{L,i}(C_j)$ and $d > d_{L,i}(C_j)$, then only these equilibria exist.

The proof is in the appendix. We do not investigate in this paper how banks could coordinate on equilibrium E2.[9] In the United Kingdom, settlement banks in CHAPS set themselves throughput

[9] There are a number of theoretical arguments against the choice of a Pareto-dominated equilibrium such as E1 (for example, pre-play communication).

targets: for example, on average, 50% of the value of all payments should be submitted by noon.[10] A committee consisting of the CHAPS settlement banks and the Bank of England in its role as overseer of payment systems monitor how well these targets are met. Because these throughput guidelines are met in CHAPS, equilibrium E2 is the better description of settlement banks' behaviour in CHAPS than equilibrium E1: in the absence of operational problems, banks transmit about half of their payments (in terms of value) in the morning. E2 predicts exactly that. If there is an operational shock, banks tend to stop sending to the stricken bank in the morning (see the estimation results in the following section), while they react much less to such a shock in the afternoon. Again, this is in line with E2's predictions.

Equilibrium E2 is efficient among symmetric[11] equilibria: given that each bank posts one unit of liquidity, payments are settled as early as possible, minimising expected costs. The underlying reason is that delay costs are higher than the cost of liquidity. Admittedly, this basic prediction has already been made in Bech and Garrat (2003). The extensions made in our model allow to derive a more precise prediction: that even when delay costs are high, banks will stop sending payments to a stricken bank in the morning, but not in the afternoon.

A bank stops sending payments to a stricken bank when it is unsure about whether it has sufficient liquidity available to execute all remaining payments. This is, in principle, good news for systemic risk. Whether it is sufficient to contain the effects of the shock is an empirical question. Our empirical results, presented in the following section, show that indeed, healthy banks' payment behaviour to a stricken bank is sufficient to leave payment values exchanged between healthy banks unaffected by the stricken bank's operational shock.

4 Estimation

This part of the paper is organized as follows. Section 4.1 provides a brief description of the UK large-value payment system; section 4.2 describes the data; in section 4.3, we analyse the impact of an operational failure on payment flows to stricken banks.

[10] See section 4.1 for more on CHAPS.

[11] Notice that we have not investigated asymmetric equilibria. These may indeed exist. For sufficiently small likelihoods of operational shocks, $d''_{L,i}(2) = d_{L,i}(1), d'_{L,i}(2) > d'_{L,i}(1)$, and $\gamma_{L,i}(2) = \gamma_{L,i}(1)$. Thus, $d_{L,i}(1) = \max\left\{d''_{L,i}(1), d'_{L,i}(1)\right\}$ may rise when one player decides to post two units of collateral, reducing that player's incentive to post any collateral at all. (See the appendix for the notation.)

4.1 A Description of CHAPS Sterling

CHAPS is the United Kingdom's high-value payment system, providing Real-Time Gross Settlement of credit transfers. CHAPS started operating in 1984 as a nationwide, electronic inter-bank system for sending irrevocable, guaranteed and unconditional sterling credit transfers from one settlement member to another for same-day value. In April 1996, it was developed into an RTGS system. It now handles nearly all large-value same-day sterling payments between banks, other than those relating specifically to the settlement of securities transactions.

Fifteen banks are direct member of the system which on business days operates between 06:00 and 16:00. In 2006 average daily volumes and values amounted to 131,000 payments and £ 231bn. Payment flows are highly concentrated. The 5 biggest banks account for over 80 per cent of both volume and value. A Memorandum of Understanding between the Bank of England (BoE) and CHAPSco sets out the respective roles and responsibilities of the BoE, CHAPSCo and the members in the operation of the CHAPS services. The BoE operational responsibilities include amongst others: ensure that settlement facilities are available for 99.95% of the operating day on average over the course of the month; settle transactions within 30 seconds; process a peak day's volumes within 4 hours; inform CHAPSCo of operational problems within 5 minutes of their identification; provide at least one month's notice of planned technical changes that may affect the system functioning. Members are required to inform CHAPS (and subsequently other members) of operational problems within 15 minutes of their identification. Further, to improve the efficiency of liquidity usage by preventing any one institution from hoarding liquidity, members are required to comply with the following guidelines, measured over a calendar month. An average of 50% of value should be throughput by 12:00 and 75% by 14:30. The other role of the BoE is to supply collateralized intraday liquidity to CHAPS members. Collateralised intraday credit and incoming payments are the main sources of liquidity in CHAPS sterling; in addition, settlement banks can also use their reserve account balance to finance payment outflows.

4.2 Data

The focus is on the payment activity of the five major banks which represents 80 per cent of the activity in value. The dataset covers 8 days in 2007 when at least one of these banks was unable to send any payment during a certain time interval. Detailed information on the timing of outages and the identity of stricken banks was provided by APACS (the UK trade association for payments). Table 1 reports for each outage, the date, start time and end time. For confidentiality reasons the identity of the bank experiencing the outage is not given. Note that the eight outages differ along several dimensions: start time, length and duration. This will allow to let outage impacts to vary along these dimensions and make this exercise more informative.

Outages	Date	Start time	End time	Start time in seconds	End time in seconds	Duration	Control Days
1	March 19th	07:00	08:10	25200	29400	1:10	March 16th
2	April 27th	15:05	15:50	54300	57000	0:45	April 26th
3	May 29th	12:33	12:51	45180	46260	0:18	May 25th
4	June 1st	12:24	13:17	44640	47820	0:53	May 31th
5	June 11th	06:00	07:40	21600	27600	1:40	June 8th
6	September 3rd	06:05	08:30	21900	30600	2:25	August 31st
7	September 4th	13:14	13:30	47640	48600	0:16	August 31th
8	October 8th	06:59	07:35	25140	27300	0:36	Octobr 5th

Table 1: Outages in 2007

A second source of information is the CHAPS database that contains individual transaction data. For each payment one observes the transaction date and time, the payment value, the payer and payee.

The time between transactions is the reciprocal of the transaction rate, which is itself a proxy for volume. We are, however, interested in payment values, as funding and delay costs are presumably proportional to the value of a payment. (Our theoretical model abstracted from the difference between volume and value for simplicity.) Following Gourieroux et al (1998) value-weighted payments durations are calculated as follows. Assume that we observe on every day a sequence of payments, which are indexed by $n, n = 1, ..., N_m$ and the associated payment times $d_n(m)$. The duration between the successive ticks $n-1$ and n is simply the time that expires between two payment times,

$$\tau_n(m) = d_n(m) - d_{n-1}(m) \tag{1}$$

The weighted durations instead represent the time required by a bank to make a fixed value υ of payments. Let $\upsilon_n(m)$ denote the value paid at time $d_n(m)$. By summing up values of individual payments for a count of $N_t(m)$ payments, the cumulated payments value is obtained

$$V_t(m) = \sum_{n=1}^{N_t(m)} \upsilon_n(m) \tag{2}$$

i.e. the volume paid on day m by t. The value duration is defined

$$\tau_{val}(t, \upsilon) = \inf\left(\tau : V_{t+\tau}(m) \geqslant V_t(m) + \upsilon\right) \tag{3}$$

as the time necessary to observe an increment v of cumulated value. v is set to 1 billion pound[12].

Variables	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Outgoing durations	1	517	930	1355.7	1705	12259
Incoming durations	1	530	922	1287.6	1611	12871

Table 2. Descriptive statistics

Before durations are calculated the values of simultaneous payments are summed (by bank) and then outgoing and incoming payment values at each point in time are matched by payer. Overnight durations are ignored. After deleting this observations there are 466348 durations (observations). Table 2 reports descriptive statistics. For outgoing durations the average time between successive events is 1355.7 seconds (or about 23 minutes). The minimum duration is 1 second and the maximum duration 12259 seconds (or about 3 hours 40 minutes). The descriptive statistics for incoming durations are about similar confirming that the bulk of the activity is processed by the 5 largest banks. Figure 1 is a plot of the density for the waiting times showing in more details the events

[12] This threshold is selected because it belongs to the top one percentile.

distribution. Durations above 1 hour occur rarely.

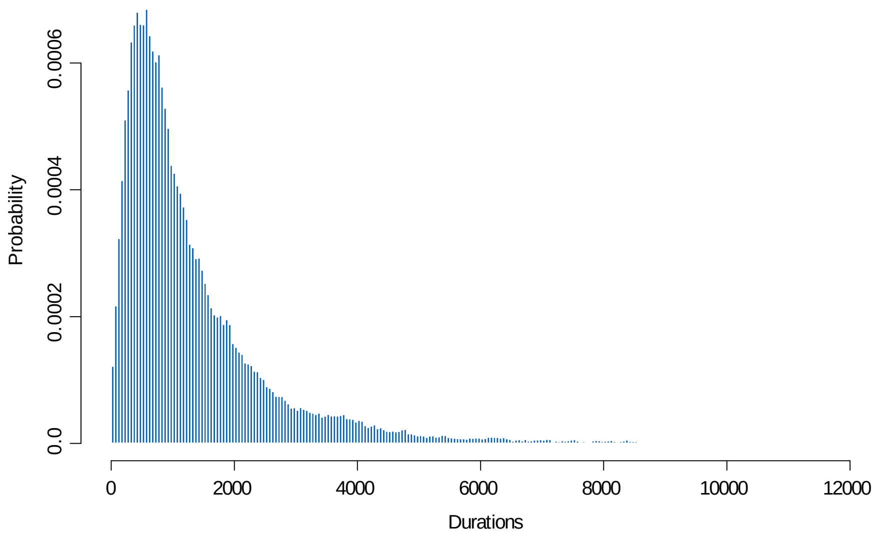

Figure 1: Density of incoming payments durations. Durations are defined as the time it takes for a bank to receive additional payments worth £1bn.

Figure 2 contains a plot of the expected time interval conditioned on the time of day that the duration begins[13]. The x-axis shows the time of the day. The y-axis shows that estimated time it takes for a bank to receive additional payments worth £1bn.[14] For example, at 8:30 (corresponding to 30,000 seconds), a bank expects to receive an additional £1bn within the next 30 minutes (corresponding to 1,800 seconds). There is a rapid increase in activity immediately after the opening and a gradual increase thereafter (with two small peaks at the throughput deadlines 12:00 (43,000

[13] The expectation is calculated using Friedman's super smoother.

[14] The estimates are derived using Friedman's super smoother.

seconds) and 14:30 (52,000 seconds).

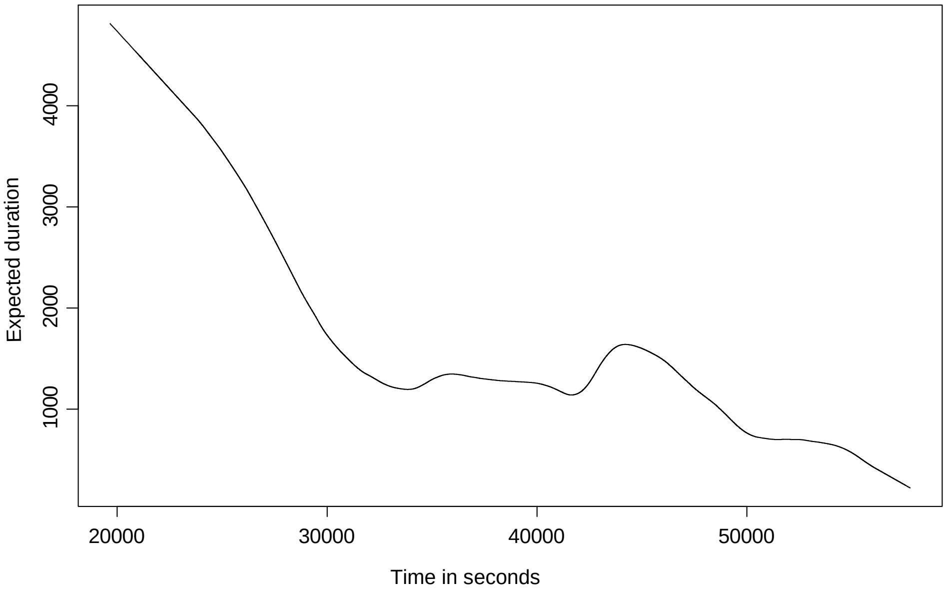

Figure 2: Daily pattern of incoming durations.

Unlike other financial data (e.g. transaction data), long durations, and likewise short durations, do not occur in clusters. The absence of duration clustering is also visible in the autocorrelation function (ACF) and partial ACF plotted in Figure 3. Indeed, autocorrelation shows up in a slowly decreasing autocorrelation function that starts at a high value and the partial autocorrelations are small in magnitude and not significant statistically.

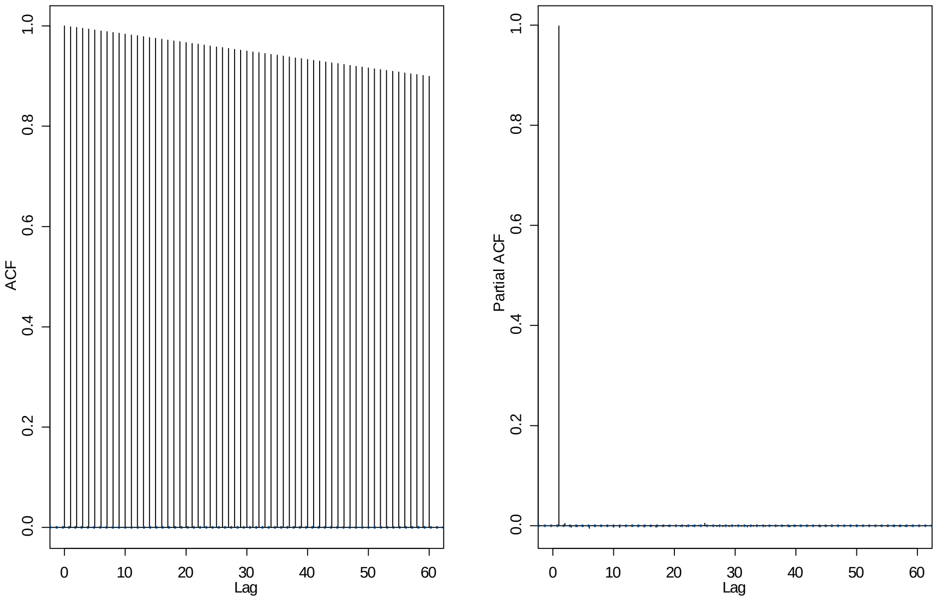

Figure 3: Autocorrelation and Partial Autocorrelation functions of incoming durations.

4.3 The Impact of Outages on Payment Flows to Stricken Banks

In this section, we estimate average differences in payment flows to stricken banks between days when they experience an outage and days when they do not experience any outage. In particular, we estimate reactions of the value-weighted duration of incoming payments: that is, the reaction of the time it takes the stricken bank to receive a certain amount of liquidity from the other banks. The higher the value-weighted duration, the less liquidity the bank receives. We also analyse how banks' reaction changed in the second half of 2007, which we associate with times of greater uncertainty in the market.

4.3.1 Empirical Specifications

Assume a bank experienced an outage started at time T_s and ended at time T_e on day d. Owing to the absence of duration clustering the following (non-dynamic) semi-parametric specification is proposed to assess changes in the intensity of a bank's incoming payment flow before, during and after outages it experienced[15]

$$I_i^b = c + Outage_{bd} * before + Outage_{bd} * during + Outage_{bd} * after + f_1\left(t_i\right) + \gamma_b \quad (4)$$

where I_i^b is the standardized value-weighted duration associated with the i^{th} incoming payments to bank b from any other banks. I_i^b is standardized by dividing the actual duration for bank b at transaction i on day d by the average duration for bank b on that day. This way the mean of I_i^b is 1 for all banks and the estimates can be interpreted as percentage deviations from the mean.

Outage is a dummy taking value one if bank b experiences an outage on day d; c is a constant; $during$ $(after/before)$ is a dummy that takes value one if $T_s \leq t_i \leq T_e$ $(t_i \succ T_e/t_i \prec T_s)$; γ_b is a set of bank fixed effects. $f_1\left(t_i\right)$ is an unspecified function to be estimated that controls for time-specific effects. γ_b are banks fixed effects, c is a constant, and t_i is the time at which duration i starts. Hence, this specification exploits variations within bank and across days. This specification allows to assess the impact of the average outage.

The next specification is used to analyse intra-outage dynamics

$$I_i^b = c + f_2\left(N_{id}\right) + f_3\left(t_i\right) + \gamma_b \quad (5)$$

where $N_{id} = t_i - T_s$ is the number of seconds elapsed since an outage started for intra-outage transactions (i.e. if $t_i < T_e$) and zero otherwise. Hence, $f_2\left(N_{id}\right)$ measures how the incoming duration to the stricken bank depends on the 'age' of the outage, that is, the time that has expired since the outage started. $f_3\left(t_i\right)$ is again included to control for time-of-the-day effects.

Last, the effect of outages is allowed to vary depending on the time of the day outages start

$$I_i^b = c + Outage_{bd} * before + OdMorning_i + OdAfternoon_i + Outage_{bd} * after + f_4\left(t_i\right) + \gamma_b \quad (6)$$

where $OdMorning$ $(OdAfternoon)$ is a dummy that takes value one if bank b experienced an outage at time t_i and t_i is a pre-12 pm (post 12pm) time.

The control days are taken as days when no bank experiences an outage. We take the closest previous working day as a control day for an outage day: inflows to a bank on a day and hours when it experiences an outage is compared to its inflow at the same hours on the closest previous working day when no bank experiences an outage.

[15] Engle and Russell (1998) develop a model of intertemporally correlated event arrival times applied to IBM transaction data.

4.3.2 Results

Table 3 column (1) reports the results of estimating equation (4). The duration it takes for a stricken banks to receive an additional billion pound of customer payments from other banks rises by about 60 per cent during an outage and recovers slowly post-outage.

Figure 4 plots the intra-outage dynamic, that is, function f_2 in equation (5), $I_i^b = c + f_2(N_{id}) + f_3(t_i) + \gamma_b$. The duration rises by up to 100% during outages. This peak is reached about 2500 seconds (or 40 minutes) after outages start and declines slightly thereafter to stabilize at 60% until the outage ends. This non-linearity may be explained by the trade-off banks face between paying immediately and incurring a liquidity cost (e.g. by having to raise additional liquidity in the inter-bank market) and delay costs. These results are qualitatively similar when one considers in isolation the longest outage (not reported).

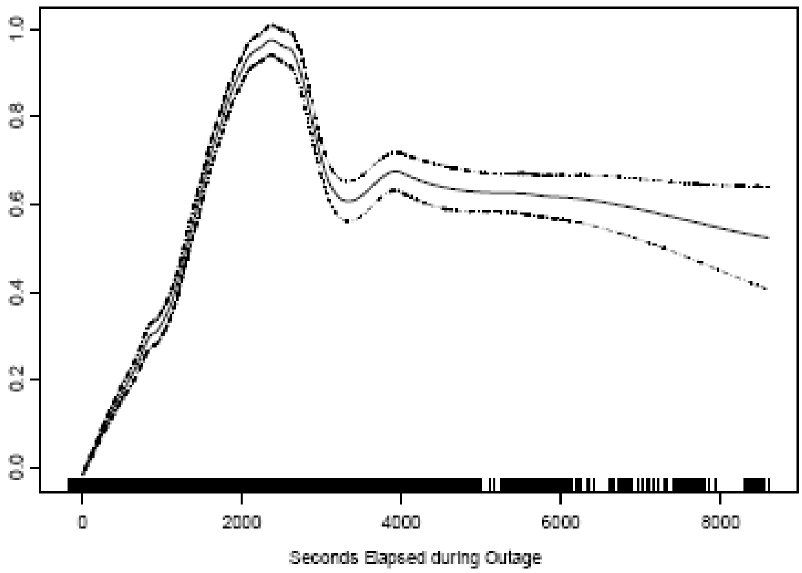

Figure 4: Intra-outage dynamics. This figure reports how incoming durations to the stricken bank evolve since the start of the respective outage (ie, a non-parametric estimate of function f_2 in equation 5)

Table 3 column (2) reports an estimation of equation (6). The results are striking. The duration of payment inflows to a stricken bank rises by 150% during outages occurring before 12 pm. It rises by only 20% during outages that occur in the afternoon. This result was to be expected as banks have more room to delay payments early in the day than late in the day. The difference between morning and afternoon outages effects falls when we exclude the two longest and two shortest outages (this makes morning and afternoon outages more comparable, column 3), but the increase in the duration is still about twice as large in the morning. (The difference remains statistically significant.) In column (4) the results are robust to ignoring outages that occurred during the sub-prime crisis. Column (5) compares the effect of the shortest two outages that occurred during the credit crunch (outages 7 and 8) to the two shortest outages pre-credit crunch (outages 2 and 3). One may conclude that *short* outages cause banks to hoard payments during the liquidity crunch (post August 9th 2007) but not in tranquil times. In column (7) we compare crisis afternoon and crisis morning outages and the results confirm a higher reaction to outages during the financial crisis . (Our model does not give an unambiguous prediction regarding behaviour in stressed and normal times, see the discussion in section 6.5.) Finally, column (7) reports an estimate of the effect of outages on activity between healthy banks. In order to derive this result incoming durations of healthy banks were calculated excluding payments from and to a stricken bank. The result indicate that outages do not produce negative externalities. None of the coefficients is significant statically implying that activity among healthy banks is unaltered during an outage.[16]

[16] Note that another estimation reveals that outages cause a statistically significant 20 per cent decline in incoming payment flows of healthy banks, abstracting from payments from the stricken bank.

Dependent variable: Incoming duration	(1)	(2)	(3)(a)	4(b)	(5)(c)	6(d)	(7)(e)
Before Outage	0.058	0.068	0.076	0.071	0.025	0.065	-0.161*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.124)
During Outage	0.596						0.057*
	(0.008)						(0.124)
After Outage	-0.08	-0.089	-0.12	-0.115	-0.07	-0.055	-0.074*
	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.005)	(0.124)
During Morning Outage		1.516	0.476	1.242		2.177	
		(0.015)	(0.02)	(0.018)		(0.025)	
During Afternoon Outage		0.206	0.206	0.163		0.450	
		(0.01)	(0.01)	(0.011)		(0.034)	
During Crisis Outage					0.516		
					(0.028)		
During non-Crisis Outage					-0.025		
					(0.012)		
Time effects	x	x	x	x	x	x	x
Bank fixed effects	x	x	x	x	x	x	x
Nobr. Observations	149811	149811	149811	105505	73436	44306	548492
R-squared	0.48	0.5	0.47	0.49	0.5	0.52	0.38
F-statistic p-value	0	0	0	0	0	0	0

Table 3: Banks' payments behaviour. We report alternative estimates of the following regression:

$$I_i^b = c + Outage_{bd} * before + Outage_{bd} * during + Outage_{bd} * after + f_1(t_i) + \gamma_b$$

where I_i^b is the standardized value-weighted duration associated with the i^{th} incoming payments to bank b from any other banks. $Outage_{bd} * before$, $Outage_{bd} * during, Outage_{bd} * after$ are dummy variables taking value 1 on days when bank b experiences an outage. $f_1(t_i)$ is a time of day effect function to be estimated, so t_i is the time at which the i^{th} duration starts. γ_b is a set of bank fixed effects and c a constant. The various specifications allow or not for specific morning and afternoon effects during outages.

All coefficients are significant at the 1% level, except estimates marked by (*).

(a) Ignores the two longest and the two shortest outages. (b) Ignoring outages occurring post August 9th (i.e. post crisis)

(c) Compares outages 7 and 8 to outages 2 and 3. (d) Only post August 9th outages i.e. crisis outages

(e) Estimates on durations calculated on all payments except from and to a stricken bank.

5 Conclusion

The evidence in this paper indicates that banks react to large operational outages by ceasing to make payments to stricken banks in the worst case scenario. In line with the prediction of our model, the reaction to outages is stronger in the morning than to those in the afternoon. The peak of the reduction of flows to the stricken bank occurs no later than one hour into the outage. Presumably delay costs become too large afterwards, encouraging banks to make some payments to the stricken bank. The fact that banks initially stop making payments to stricken banks reduces systemic risk: the stricken bank does not become a liquidity sink, and liquidity remains available to settle outstanding payments between healthy banks. Indeed, we show that the value of payment flows between healthy banks remain virtually unchanged during an outage.

References

[1] Angelini P. (1998), "An analysis of competitive externalities in gross settlement systems", Journal of Banking and Finance, 22(1):1-18

[2] Armentier O., Arnold J. and McAndrews J. (2007), "Changes in the timing distribution of Fedwire funds transfers", FRBNY mimeo

[3] Bech M. and Garratt R. (2003, "The intraday liquidity management game", Journal of Economic Theory, 109(2):198-210

[4] Becher C., Galbiati M. and Tudela M. (2007), "The timing and funding of CHAPS sterling payments", Bank of England mimeo

[5] Engel R., Russell J. (1998), "Autoregressive Conditional Duration: a new model for irregularly spaced transaction data", Econometrica, Vol 66, N0. 5 pp 1127-1162

[6] Gourieroux C., Jasiak J., Le Fol G. (1999), "Intra-day market activity", Journal of Financial Markets 2, 193-226

[7] Khan C., McAndrews J. and Roberds W.(2003), "Settlement risk under gross and net settlement", Journal of Money Credit and

[8] Banking, 35(4):591-608

[9] McAndrews J. and Rajan S. (2000), "The timing and funding of Fedwire Funds transfers", FRBNY Economic Policy Review, 6(2):17-32

[10] McAndrews J. and Potter S. (2002), "Liquidity effects of the events of September 11, 2001", FRBNY Economic Policy Review 8(2):59-79

[11] Mills D. and Nesmith T. (2006), "Risk and concentration in payment and securities settlement systems", mimeo Federal Reserve Bank of New York

6 Appendix

6.1 Proof of the main proposition

The proof proceeds by backwards induction. Equilibrium play in the afternoon and evening is proven in section 6.2. Lemma 1 shows that $d > d''_{L,i}(C_j)$ ensures that $p_i^M = 1$ only if $p_j^M = 1$, implying that if $C_i = C_j = 1$, there is an equilibrium in which no player pays in the morning (E1), and another one in which payments are exchanged only if neither player is hit by an operational shock (E2). Lemma 2 shows that $d > d'_{L,i}(C_j)$ ensures $C_i \geq 1$. Definition 1 defines $d_{L,i}(C_j)$:

Definition 1 $d_{L,i}(C_j) = \max\left\{d'_{L,i}(C_j), d''_{L,i}(C_j)\right\}$.

Finally, Lemma 3 shows that $\gamma > \gamma_{L,i}(C_1)$ ensures $C_i \leq 1$. The non-existence of other equilibria is simply given by assuming that the conditions hold for all opponent actions (all C_j).

The following sections guide the reader through the proofs. Given that there are six independent operational shocks (one for each player in each period morning, afternoon and evening), the detailed proofs are cumbersome (but straightforward), as optimal behaviour has to be computed, at least in principle, for every possible realisation. We therefore decided to only give examples, and to offer to make the complete calculations available upon request.

6.2 Equilibrium play in the afternoon and evening

In the afternoon, each player knows all his outstanding payment instructions and his available liquidity. There is no gain from hoarding liquidity but a positive cost because there is a risk that i will be unable to make the delayed payment in the evening. Priority is always given to the urgent payment because the cost of delay and the cost of technical default for the urgent instruction are higher than for the normal instruction ($d > 0$ and $f_u > f_n$). Technical default is, by assumption, sufficiently expensive for i to always attempt to raise additional liquidity if necessary. (Assuming that the attemt to raise additional liquidity succeeds with probability 1/2, this holds if $-\left(\frac{1}{2}f_{\cdot} + \frac{1}{2}\gamma\right) > -f_{\cdot}$, which is implied by $-\gamma > -f_n, -f_u$.). Thus, in the afternoon, all available liquidity is used to execute outstanding payment instructions unless the player is hit by an operational shock. Normal instructions are only executed after urgent instructions have been executed. In the evening, all remaining instructions are executed subject to available liquidity unless the player is hit by an operational shock. If liquidity is insufficient, and $s_i^E = 0$, an attempt is made to raise additional liquidity.

Interestingly, player i's preferred actions are independent of opponent play. This follows because there is no intra-period netting of payments: independently of how many instructions the opponent executes, i needs to have all liquidity necessary to make his payments at the start of the corresponding period. Also, because the costs for the liquidity available in the afternoon have already been incurred at the start of the day, there is no possibility to 'save' liquidity costs by waiting for incoming payments. But in expectation, even the delay of normal payments is costly because there is a risk that i might suffer an operational outage in the evening, and hence fail to execute the outstanding payment instructions.

6.3 Equilibrium play in the morning

The key result in this section is that if $C_i = 1$, then there is a threshold $d''_{L,i}(C_j)$ for the delay cost of the urgent action above which i withholds payments to j in the morning if j experiences an operational outage. (When $C_i = 0$, then $p_i^M = 0$ is the only possible action, and when $C_i = 2$, then i has sufficient liquidity to execute even the urgent payment transaction without having to rely on incoming liquidity, and executes in equilibrium each transaction as it arrives.) The proof proceeds by explictly computing the expected payoffs of the actions $p_i^M \in \{0, 1\}$ over all possible realisations of the random variables s_i^t, and given the opponent's choice p_j^M, and the collateral posting choices made by both players. For example, if $C_i = 1$, $C_j = 0$, then i's expected payoff $u_i^M\left(C_i, C_j, p_i^M, p_j^M\right)$ from playing $p_i^M = 0$ is

$$\begin{aligned} E_{\left(s_1^A, s_2^A, s_1^E, s_2^E\right)} u_i^M\left(1, 0, 0, p_j^M\right) &= -\varepsilon\left(1-\varepsilon\right)\left(-v_i f_n\right) + \left(1-\varepsilon\right)^2\left(-v_i\left(\frac{1}{2}f_n + \frac{1}{2}\gamma\right)\right) \\ &\quad + \varepsilon^2\left(-f_n - v_i f_u\right) + \varepsilon\left(1-\varepsilon\right)\left(-v_i\left(d + \frac{1}{2}f_n + \frac{1}{2}\gamma\right)\right) \end{aligned}$$

To see this, notice that $C_j = 1$ implies $p_j^M = 0$. Then $l_i^A = 1$, $l_j^A = 0$, $p_j^A = 0$, $l_i^E = 1 - p_i^A$ and i's payoffs are independent of s_j^A because the earliest time j can pay is in the evening. Payoffs now depend on the realisation of the random variables $\left(s_1^A, s_2^A, s_1^E, s_2^E\right)$: If $s_i^A = 0$, $s_i^E = 1$, i's expected payoff is $-v_i f_n$; if $s_i^A = 0$, $s_i^E = 0$, then i's expected payoff is $-v_i\left(\frac{1}{2}f_n + \frac{1}{2}\gamma\right)$; if $s_i^A = 1$, $s_i^E = 1$, then i's expected payoff is $-f_n - v_i f_u$; and if $s_i^A = 1$, $s_i^E = 0$, then $l_i^E = 1$, and i's expected payoff is $-v_i\left(d+\frac{1}{2}f_n+\frac{1}{2}\gamma\right)$. Correspondingly, expected payoffs can be derived for the other cases; these calculations are omitted here but available from the authors.

$d''_{L,i}(C_j)$ is then determined by solving $Eu_i^M(1,C_j,0,0)-Eu_i^M(1,C_j,1,0)>0$ for d, where the expectations are again over $\left(s_1^A,s_2^A,s_1^E,s_2^E\right)$. It is given by definition 2:

Definition 2

$$d''_{L,i}(C_j)=\begin{cases} d''_{L,i}(1)+(\gamma+f_n)\frac{1-\varepsilon_j}{2-2\varepsilon_i-\varepsilon_j} & \text{if} \quad C_j=0 \\ \frac{2\varepsilon_i^2 f_n-v_i(1-\varepsilon_i)(2\varepsilon_i+\varepsilon_j)(f_u-f_n)}{v_i(1-\varepsilon_i)(2-2\varepsilon_i-\varepsilon_j)} & \text{if} \quad C_j\in\{1,2\} \end{cases}$$

Notice that $d''_{L,i}(1)>d''_{L,i}(0)$. This is because there is an additional incentive to pay early when $C_j=0$: if i pays in the morning, he might be able to re-use this unit of liquidity in the evening.

It is instructive to have a closer look at $d''_{L,i}$ for the special case in which there is hardly any risk of operational failure ($\varepsilon_i,\varepsilon_j\to 0$). Then

$$d''_{L,i}(C_j)_{\varepsilon_i=\varepsilon_j=0}\to\begin{cases}\frac{1}{2}(\gamma+f_n) & \text{if} \quad C_j=0\\ 0 & \text{if} \quad C_j\in\{1,2\}\end{cases}$$

Clearly, near-absence of operational shocks means that the incentive to make payments early is (nearly) lost when the opponent j has posted some collateral: j will use this liquidity if not in the morning, then in the afternoon period to make a payment; i can use this to make a second payment in the evening. Thus, for all $d\geq 0$, i prefers to reserve the payment for the (possible) urgent payment instruction. In contrast, when $C_j=0$, there is a benefit from paying in the morning: the opponent can then use this liquidity in the afternoon, and i can re-use it in the evening. If he receives an urgent payment instruction (probability v_i), i thus saves the need to attempt to raise additional collateral in the evening for the remaining normal instruction (which, if successful, costs γ, whereas the cost of failure is f_n). If $v_i d>v_i\left(\frac{1}{2}\gamma+\frac{1}{2}f_n\right)$, these costs are dominated by the cost of delaying the urgent payment, and i prefers to save his liquidity for the afternoon. Thus, $d''_{L,i}=\frac{1}{2}\gamma+\frac{1}{2}f_n$ if $C_j=0$.

Lemma 1 *If $C_i=1$, and he is not hit by an operational shock ($s_i^M=0$), his optimal payments behaviour depend on the opponent's payment behaviour, and on the cost d of delaying the execution of the urgent payment instruction:*

$$p_i^M=p_{n,i}^M=\left(1-s_i^M\right)\cdot\begin{cases}1 & \text{if } p_j^M=1, \text{ or if } p_j^M=0 \text{ and } d\leq d''_{L,i}(C_j)\\ 0 & \text{if } p_j^M=0 \text{ and } d>d''_{L,i}(C_j)\end{cases}$$

Proof 1 *Consider first the case that $p_j^M=1$. If $p_i^M=1$, then $l_i^A=C_i-p_i^M+p_j^M=1$, sufficient to execute the urgent transaction immediately should it arrive. Thus, there is no benefit from delaying* *the execution of the normal payment, but an expected cost, given that i might suffer an operational outage in the subsequent periods.*

Now suppose that $p_j^M = 0$*. From the definition of* $d''_{L,i}(C_j)$*,* $Eu_i^M(1, C_j, 0, 0) - Eu_i^M(1, C_j, 1, 0) > 0$ *if only if* $d < d''_{L,i}(C_j)$*, implying that it is optimal for i to withhold the morning payment if and only if* $d < d''_{L,i}(C_j)$*.*

6.4 Optimal collateral posting at the start of the day

Lemma 2 provides an intuitive condition under which $C_i = 0$ is dominated. Lemma 3 then investigates under which conditions $C_i = 1$ is preferred over $C_i = 2$. We then argue that these two conditions are independent.

Lemma 2 *For sufficiently high* $d > d'_{L,i}(C_j)$*, bank i prefers posting one or two units of collateral over posting no collateral.*

This result should be intuitive: if the costs of delaying the urgent payment are sufficiently high (and the instruction sufficiently likely to arrive), then i does not want to rely on incoming liquidity in the first round to finance the urgent payment instruction. The proof first computes i's expected payoff $Eu_i(C_i, C_j)$ from posting C_i units of collateral at the start of the day, given optimal play by both players in the subsequent rounds (see lemma L11). The expectation is over realisations of the operational shock $\left(s_1^M, s_2^M, s_1^A, s_2^A, s_1^E, s_2^E\right)$. $d'_{L,i}(C_j)$ is then determined by solving $Eu_i(1, C_j) - Eu_i(0, C_j) > 0$ for d. It is omitted here but avaialable from the authors.[17]

To illustrate the results with a special case, assume that there is hardly any risk of operational failure ($\varepsilon_i, \varepsilon_j \to 0$), and $C_j = 1$. Then

$$d'_{L,i}(1) = \gamma / v_i - \frac{1}{2}(\gamma + f_n)$$

When $v_i \to 0$, it is always better to post no collateral: When the likelihood of operational shocks is virtually zero, i can rely on incoming liquidity. This is the basic prisonner's dilemma that Bech and Garrat (2003) identified. Assume instead that i is certain to obtain an urgent payment instruction ($v_i = 1$). If $C_j = 1$, i can only rely on one unit of incoming liquidity. If $C_i = 0$, j does not pay in the morning (by lemma 1), and i has to delay the execution of the urgent payment transaction in the afternoon (cost: d). In addition, he has to raise the remaining unit of liquidity in the evening (cost: $(\gamma + f_n)/2$). If these costs exceed the cost γ of raising liquidity in the morning, that is, if

[17] We do not investigate the relation between d' and d'', but state the result in our proposition for sufficiently high $d > \max\{d', d''\}$.

$d + \frac{1}{2}(\gamma + f_n) > \gamma$, then i prefers to post one unit of collateral. If $v_i > 0$, then $d + \frac{1}{2}(\gamma + f_n)$ only has to be paid when i receives an urgent payment instruction, and i only prefers to post one unit of collateral in the morning if $v\left(d + \frac{1}{2}(\gamma + f_n)\right) > \gamma$, equivalent to the definition given above.

Lemma 3 investigates the decision between posting one and two units of collateral at the start of the day. Unsurprisingly, for sufficiently high costs of liquidity, posting two units is dominated.

Lemma 3 *If $\gamma > \gamma_{L,i}$, and $d > d''_{L,i}$, then i prefers $C_i = 1$ over $C_i = 2$.*

$$\gamma_{L,i}(C_j) = (1-\varepsilon_i) f_n \cdot \begin{cases} \left(v_i(1-2\varepsilon_i) + 2\varepsilon_i^2\right) / \left(2 - v_i(1-\varepsilon_i)\right) & \text{if} \quad C_j = 0 \\ \left(2v_i\varepsilon_i + v_i\varepsilon_j + 2\varepsilon_i^2\right) / \left(2 - v_i\varepsilon_j(1-\varepsilon_i)\right) & \text{if} \quad C_j = 1 \\ \varepsilon_j\left(2v_i\varepsilon_i + v_i\varepsilon_j + 2\varepsilon_i^2\right) / \left(2 - v_i\varepsilon_j^2(1-\varepsilon_i)\right) & \text{if} \quad C_j = 2 \end{cases}$$

The proof first computes i's expected payoff $Eu_i(C_i, C_j)$ from posting C_i units of collateral at the start of the day, given optimal play by both players in the subsequent rounds (see lemma L11). The expectation is over realisations of the operational shock $\left(s_1^M, s_2^M, s_1^A, s_2^A, s_1^E, s_2^E\right)$. $\gamma_{L,i}(C_j)$ is then determined by solving $Eu_i(1, C_j) - Eu_i(2, C_j) > 0$ for γ. For example, in the case of $C_j = 0$, then $p_j = 0$, so following lemma 1, $p_i^M = 0$, and $Eu_i(1,0) = -\gamma + E_{\left(s_1^A, s_2^A, s_1^E, s_2^E\right)} u_i^M(1,0,0,0)$ which was given above. If $C_i = 2$, i is independent of incoming payments, and executes all instructions immediately. Then

$$Eu_i(2,0) = -2\gamma + \varepsilon\left((1-\varepsilon)(-v_i d) + \varepsilon(-v_i f_u)\right) + \varepsilon^3(-f_n)$$

Regarding the first term, recall that $C_i = 2$ costs 2γ. Regarding the second, $\varepsilon(1-\varepsilon)(-v_i d)$, notice that i suffers a delay cost $-d$ from the execution of the urgent transaction only if he obtains such a transaction (probability v_i), and if he is unable to execute it immediately in the afternoon (probability ε) but able to execute it in the evening (probability $(1-\varepsilon)$). Correspondingly, he fails to execute the urgent transaction if he is hit by an operational outage in both the afternoon and the evening, resulting in an expected cost given by $\varepsilon^2 v_i f_u$. Finally, there is a chance that i is unable to execute the normal transaction if he is hit by three successive operational outages in the morning, afternoon and evening, resulting in a cost of $\varepsilon^3(-f_n)$. The proof for the other levels of C_j proceeds correspondingly and is available from the authors.

Again, it is instructive to look at the special case of very unlikely operational outages ($\varepsilon_i, \varepsilon_j \to 0$). Then

$$\gamma_{L,i}(C_j) = f_n \cdot \begin{cases} v_i / (2 - v_i) & \text{if} \quad C_j = 0 \\ 0 & \text{if} \quad C_j \in \{1,2\} \end{cases}$$

If there is hardly any risk of operational failure, there is no benefit from posting two units of liquidity if $C_j \in \{1,2\}$ because i can rely on incoming funds in the morning and/or in the afternoon to make a second payment. In contrast, if the opponent posted no liquidity, $p_j^M = 0$, so $p_i^M = 0$ as well because $d > d''_{L,i}$, and j will not be able to pay before the evening. Thus, if $C_i = 1$, and i obtains an urgent payment instruction (probability v_i), then i has to attempt to raise an additional unit of liquidity in the evening. If $\gamma > \frac{1}{2}v_i(\gamma + f_n)$, the cost of posting this unit of collateral at the start of the day exceeds the expected costs of attempting to raise it in the evening (which fails with probability 1/2 such that i cannot execute the remaining normal payment and incurs a cost of f_n). An equivalent expression of this inequality is $\gamma > f_n v_i / (2 - v_i)$.

The reader will have noticed that $\gamma_{L,i}(C_j)$ is independent of d. This may, at first sight, be surprising, and is an important property: we state in our main proposition that $C_i = 1$ is optimal for sufficiently high delay costs d (making posting more collateral more attractive), and sufficiently high costs of collateral γ (making posting less collateral more attractive), so it is important to show that these two conditions are independent to ensure that such (d, γ) indeed exist. The key point is that $l_i^A \geq 1$ if $C_i \in \{1, 2\}$ and $d > d_{L,i}$. Consequently, liquidity is always available to make the urgent payment. This is obvious when $C_i = 2$. To see that this also holds for $C_i = 1$, recall from lemma 1 that in this case, $p_i^M = 0$ only if $p_j^M = 0$, in which case $l_i^A = C_i = 1$. Conversely, $p_i^M = 1$ only if $p_j^M = 1$, in which case $l_i^A = C_i - p_i^M + p_j^M = 1$ as well. Consequently, if i obtains an urgent payment instruction ($v_i = 1$), then i only incurs the delay cost d if i is struck by operational problems in the afternoon ($s_i^A = 1$). But operational shocks are independent of start-of-day liquidity postings. Thus, given $C_i \in \{1, 2\}$ and $d > d''_{L,i}$, the expected delay cost is independent of of C_i.

6.5 Reaction to operational shocks in times of stress

Should we expect settlement banks' payments behaviour to chance in times of stress? In our sample period, times of stress were characterised by heightened uncertainty about a bank's ability to refinance unexpectedly high payment outflows by borrowing in the interbank market. Higher uncertainty led to a higher demand for collateral eligible for central bank credit operations, driving up the price of collateral. In a model of risk-neutral banks, this uncertainty is presumably best approximated by an increase in the likelihood that a bank may receive an urgent payment instruction (that is, the bank becomes more likely to need more collateral), and an increase in the opportunity costs of collateral: that is, by an increase in v_i, v_j, and γ. The costs of delaying urgent payments, and, *a fortiori*, of failing to make a payment, presumably also rose, corresponding to an increase in d, f_n and f_u. In reality, these costs of failing to execute a payment instruction corresponds to the cost of accessing the central bank's standing facilities to finance the outstanding payment. In the second half of 2007, banks were indeed reluctant to make use of standing facilities in the UK, fearing that this might be interpreted as a sign of solvency problems.

Importantly, if parameters are such that the conditions for the existence of E2 continue to hold, and if operational shocks do not hit either bank, then we should expect both banks to continue to make their payment in the morning. This is, indeed, what we observed in CHAPS in the second half of 2007. As the empirical results show, however, banks' reaction to operational shocks was on average stronger than in normal times. This could be explained by banks' heterogeneity. For some, the likelihood of receiving an urgent payment instruction may have been too small to ensure that $d > d_{L,i}(1)$, such that they made their morning payment independently of their opponent's ability to send payments. The remaining question is whether E2 continues to exist when v_i, v_j, γ, d, f_n and f_u all increase. One can show that both sides of the two conditions, $\gamma > \gamma_{L,i}(1)$ and $d > d_{L,i}(1)$, may increase as a result. We find it difficult to judge which effect will dominate - after all, these variables are exogenous in our model. Hence our careful conclusion that the results do not contradict the predictions of the model.