Simulation analyses and stress testing of payment networks

Scientific monographs E:42 · 2009

Harry Leinonen (ed.)

Simulation analyses and stress testing of payment networks

Proceedings from the Bank of Finland Payment and Settlement System Seminars 2007–2008

SUOMEN PANKKI · FINLANDS BANK

EUROJÄRJESTELMÄ
EUROSYSTEMET

Scientific monographs
E:42 · 2009

Harry Leinonen (ed.)

Simulation analyses and stress testing of payment networks

Proceedings from the Bank of Finland Payment and Settlement System Seminars 2007–2008

SUOMEN PANKKI · FINLANDS BANK
EUROJÄRJESTELMÄ
EUROSYSTEMET

Scientific monographs
E:42 · 2009

The views expressed in this study are those of the authors and do not necessarily reflect the views of the Bank of Finland.

ISBN 978-952-462-512-8
ISSN 1798-1077
(print)

ISBN 978-952-462-513-5
ISSN 1798-1085
(online)

Multiprint Ltd
Helsinki 2009

Abstract

This publication consists of ten separate studies on payment and settlement systems employing simulation techniques. Most of these were carried out using the payment and settlement system simulator BoF-PSS2 provided by the Bank of Finland. The preliminary versions were presented at the annual simulator seminars arranged by the Bank in 2007 and 2008. The main focus of the analyses is on continuity arrangements, operational stability, liquidity requirements, liquidity economising, gridlock resolution, transaction queuing arrangements, network features and network topologies. The studies examine systems in several countries and cover different kinds of payment systems and regimes.

Keywords: simulation, payment and settlement system, liquidity, gridlock, systemic risk, network topology

JEL classification numbers: C15, C61, D53, G10, G18, G28

Tiivistelmä

Tämä julkaisu koostuu kymmenestä erillisestä maksu- ja selvitysjärjestelmää koskevasta tutkimuksesta, jotka on tehty simulointimenetelmiä käyttäen. Useimmissa näistä tutkimuksista on käytetty Suomen Pankin maksu- ja selvitysjärjestelmäsimulaattoria BoF-PSS2. Alustavat versiot tutkimuksista on esitelty Suomen Pankin järjestämien vuosittaisten simulaattoriseminaarien yhteydessä vuonna 2007 tai 2008. Pääpaino tutkimuksissa on ollut poikkeustilanteiden ratkaisuissa, toiminnallisessa vakaudessa, likviditeettitarpeiden selvittämisessä, likviditeetin käytön tehostamisessa, lukkiutumistilanteissa ja niiden avaamiseen liittyvissä metodeissa, tapahtumien jonotuskäytännöissä, maksuverkkojen ominaisuuksissa ja verkkotopologiassa. Tutkimukset koskevat eri maissa ja eri periaatteilla toimivia järjestelmiä.

Asiasanat: simulointi, maksu- ja selvitysjärjestelmä, likviditeetti, lukkiutumistilanne, systeemiriski, maksuverkkotopologia

JEL-luokittelu: C15, C61, D53, G10, G18, G28

Preface

Payment and settlement systems are an important and integral part of modern economies. With globalisation moving ahead, payment networks are expanding rapidly and the actors are becoming increasingly interdependent. The current financial turmoil clearly spotlights these interdependencies. Understanding the network characteristics of the expanding payment and settlement systems have become key issues for central bank oversight. The need for more international cooperation in payment network design, operations and oversight, in both normal and abnormal situations, has increased considerably. Both banks and authorities will face new types of continuity requirements in a complex network of business relationships. It is important to gain a good understanding of the systems and their interdependencies in order to increase their efficiency for all kinds of circumstances.

Payment and settlement systems have proven to be a complicated area, for which simulation techniques provide a good way to penetrate sufficiently deeply. Models can be built to closely replicate the actual operating environment and can be used for testing and observing scenarios not normally found in real operating environments. Two lessons from the current crisis are that even very extraordinary situations need to be stress-tested in simulation models and that participant interdependencies can have systemic consequences. The current crisis, at least so far, has proven that the work done during the last decade in improving the resiliency and stability of payment and settlement systems has been rewarding, as all major systems have been working without problems throughout these difficult times.

The Bank of Finland has a long tradition of economic research and economic modelling, and modern payment and settlement systems have been one of the focal areas. Research based on simulation models for payment systems was initiated around the time Finland was joining the Economic and Monetary Union, and it proved an excellent tool for studying the changing liquidity needs and system risks under the new EMU regime. Based on positive results and feedback, the Bank of Finland decided to develop a diversified simulator BoF-PSS2 especially for external use and international distribution. It was completed in spring 2004 and is available for research purposes free of charge. It is a service offer under continuous development and the most recent addition of this year has been a network analysis module. Currently, the simulator has over 70 users worldwide and on every continent. While the users are mainly central banks, in recent years the interest in the simulator has increased among academics and private in-frastructure organisations.

The simulator investment and service attracted large international interest and a variety of research at various central banks. The Bank of Finland has arranged six yearly international payment and settlement seminars and workshops during the years 2003 to 2008 and the seventh seminar will be held later in 2009. The main goals of the seminars and workshops are to stimulate simulation-based payments and securities settlement research, to share research results and experiences among the user community, and to obtain new ideas and feedback regarding simulator development needs. The presentations of the first two seminars were published in the first simulator publication (BoF publication E31:2005), and those of the third and fourth seminars were published in the second publication on seminar proceedings (BoF Publication E39:2007). The presentations of the fifth and sixth seminars are included in this publication.

I would like to thank the authors for their contributions to this publication, which, I trust, will provide a good introduction to the simulation analysis of payment and settlement systems and will stimulate further research that will enhance our understanding and improve the models and methodologies in the years ahead.

I would like to thank all other contributors, sponsors, commentators and users of the simulator for all the help they have provided during the various stages of work on and with the simulator.

At the same time I would like to acknowledge Ville Ruoppi of MSG Software, who has been continuously involved in the IT aspects of the simulator. A detailed list of acknowledgements can be found on the simulator website and in the users' manual for the simulator.

For the finalisation of the publication we are indebted to Päivi Nietosvaara for the text editing, Glenn Harma for English language revision and Teresa Magi for printing administration. We are also indebted to the editorial board of the publication, consisting of Harry Leinonen, Marianne Palva and Jouko Vilmunen.

I hope users of the simulator will continue to be active and that the simulator will attract new users and sponsors. It is a great pleasure for me to present, via this third simulation-related publication on seminar proceedings, the fruits of this continuing productive cooperation between central banks.

Helsinki, June 2009
Seppo Honkapohja

Contents

Chapter 1

Introduction

Harry Leinonen

1 Introduction

The current financial turmoil has, within a short time span, changed the focus of overseers and supervisors. It has also showed how interdependent the financial markets are. These clearly comprise the same kind of large interwoven network as the internet-web. All banks and payment systems are connected in some manner to each other in one large network. Changes in one part of this network can therefore affect other, even quite distant, parts. Local oversight will not be sufficient in such circumstances; the whole network will require attention and assessment with a special focus on the most critical parts and points of contagion.

Overseeing and supervising such a global web-network will require more cooperation among the national authorities, and perhaps even new types of international cooperation mechanisms and organisations. The key authorities for oversight cooperation – the payment system committee of G10 central banks (CPSS), the payment and settlement system committee of the Eurosystem (PSSC), and the working groups of the European Commission – will need to increase and deepen their cooperation in order to fulfil the task of overseeing the web of international payment systems. There is also a need to increase cooperation between oversight and supervision authorities.

The first payment system simulation studies in the mid and late 1990s focused on settlement conventions and hidden credit and liquidity risk in individual and separated payment systems. In these studies, payment systems were mostly seen as static systems for booking interbank fund transfers. The network aspects and characteristics of payment systems have recently received more attention, as their importance has increased. The volumes in the international financial markets and across borders have increased in all types of international systems: exchanges, CCPs and CLS. In Europe the integration developments can be seen in TARGET2 and the SEPA undertaking – which result in regional, rather than national, infrastructures. We can also see a development towards multicurrency systems servicing several national currency areas. The payment system networks become more complex, and they will require new kinds of efforts from the researchers and new facilities from the tools employed.

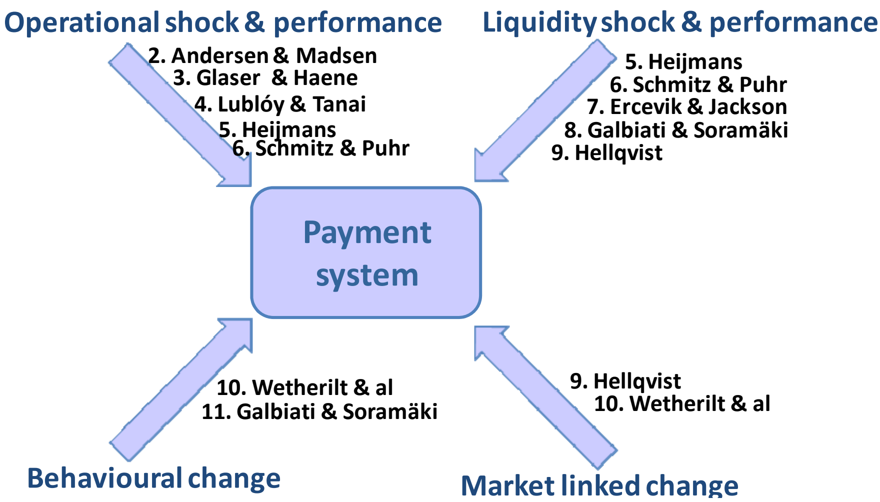
The general ICT developments support the development of data ‘crunching and mining-based’ approaches, as all necessary data are already in electronic formats and it is only necessary to bring it together within the processing of modern and efficient analysing tools.

Understanding the risks in payment and settlement systems requires studying the underlying payment flows and participant behaviour in different scenarios.

The Bank of Finland has hosted yearly simulator seminars since 2003. This is the third publication of studies presented at these seminars, covering those in 2007 and 2008. The counterparty risks and interdependencies in payment systems are of continuous concern to central banks, and stress-testing systems are becoming a general tool for overseers. The financial crises will most likely increase the interest in different kinds of stress-testing of payment systems and their participants. Abnormal situations will also change participants' behavioural patterns, resulting in new kinds of payment and liquidity flows. Payment networks will show different network characteristics when stressed.

The articles of this publication cover four main external elements, which affect the functioning of payment and settlement systems, as depicted in Figure 1.1.

Figure 1.1 **The main topics of the articles**



The chapters in the publication are ordered by starting with operational stress testing, continuing with articles on network features and liquidity issues, after which articles linking payments to intraday and overnight loan markets are presented; the final article presents an agent-based model of the payment systems. Each chapter (authors named) provides an individual stand-alone analysis, but some clearly build on earlier analyses.

Chapter 2 (Andersen and Madsen) analyses the efficiency and need for contingency measures in RTGS systems. Operational incidents in the Danish RTGS system, Kronos, are simulated in order to identify critical values for recovery time, critical participants, contingency measures and stop-sending limitations. The simulations indicate that for Kronos resumption of normal settlement within four hours and a contingency capacity between five and ten per cent of average daily transaction volumes are sufficient to significantly reduce the negative impacts of operational incidents.

Chapter 3 (Glaser and Haene) shows that in an adverse scenario an operational disruption of a major participant in the Swiss large-value payment system, Swiss Interbank Clearing (SIC), can lead to significant systemic liquidity effects. The study simulates suspension of a participant's outgoing payments, which results in liquidity accumulations for the failing participant and liquidity shortages for other participants and thereby disrupts settlement of their payments. The work highlights the importance of measures aimed at preventing operational disruptions or limiting their negative effects on other participants, such as target times for resumption of operations, contingency systems, incentivising early input of transactions, extending intraday credits, developing liquidity optimising algorithms and enhancing interbank alarm systems and crisis organisations.

Chapter 4 (Lublóy and Tanai) explores the operational resilience of the Hungarian real time gross settlement system, known as VIBER, by simulating the ability of the system to withstand operational defaults of one or two of the most important participants. The system is stressed in six different scenarios. The concentration level, availability of liquidity, back-up procedures, reactions of non-defaulted banks and structure of the money market are found to be the main factors affecting the value of unsettled transactions. Behavioural reactions of non-defaulted participants are found to be important.

Chapter 5 (Heijmans) presents an analysis of the sensitivity of the Dutch interbank payment system when stressed with liquidity shocks and operational incidents. The Dutch system is highly concentrated, with three large dominating banks and many rather small participants. A disruption of one of the large banks does not have a huge impact on the other large banks in the system. The small banks however do face more liquidity problems as a result of a disrupted (large) bank.

Chapter 6 (Schmitz and Puhr) investigates the interaction between structure and stability in payment systems. It quantifies the contagion impact of operational shocks at participants' sites in the Austrian large-value payment system, ARTIS, by means of simulations. It finds that contagion displays large variations across days and across scenarios. Based on a panel data approach, it then goes on to test whether structure can help explaining these variations. The paper captures structure by a set of network indicators. The main finding is that standard economic variables describing networks/nodes (such as liquidity, liquidity loss, value and volume of payments) can explain a substantial fraction of the variation studied. The structural indicators at the network level seem to add very little explanatory power, while those at the node level contribute somewhat to our understanding of the interaction between structure and stability in payment systems.

Chapter 7 (Ercevik and Jackson) quantifies the impact of introducing one particular type of 'hybrid design', a centralised receipt-reactive queue, on the liquidity demands faced by banks using large-value payment systems in the UK RTGS system, CHAPS. Significant liquidity savings are achievable if banks choose to enter a high proportion of their payments into the special queue. Liquidity savings are distributed unevenly; the largest users do not benefit significantly while the smaller users obtain significant savings. The level of liquidity recycling in the existing payment system is shown to be a key determinant of the impact of the hybrid design.

Chapter 8 (Galbiati and Soramäki) assesses the benefits of liquidity saving mechanisms in interbank payment systems by comparing two different payment system setups. The equilibrium choices in the two models are compared with each other and with the choices of a benevolent planner. Central queuing is found to be more efficient than decentralized queuing. The 'mechanical' advantages of a liquidity saving mechanism, at least the one assessed here, can be nullified by strategic behaviour, as there can exist 'bad' equilibria, with high liquidity usage and intensive use of the liquidity saving mechanism, and yet costs that exceed those of a system without saving mechanisms. These findings suggest that liquidity saving mechanisms are useful tools, but they may need some coordination device to ensure that banks arrive at a 'good' equilibrium.

Chapter 9 (Hellqvist) proposes an empirical method for identifying internal intraday counterparty limits of banks from payment system data. The method is tested with transaction data from Finnish RTGS system (BoF-RTGS) from years 2002–2007. The validity of the method is tested in two approaches, both based on the assumption that intraday liquidity management would reflect counterparty risk management: external market based measures and the magnitude of intraday versus overnight credit exposures. One possible observation from this study is that the difference between the uncollateralized and collateralized interbank interest rate may be positively correlated with the size of allowed intraday positions, at least during some time periods. The overall level of intraday positions was found to vary greatly compared to overnight loan positions. New developments in LVPSs enable more advanced liquidity management within the centralized payment system.

Chapter 10 (Wetherilt, Zimmerman and Soramäki) analyses the unsecured overnight market in the UK as a network of relationships and examines how it has changed during the period of market turmoil. Using established network techniques, strong evidence is found for the existence of a ‘core’ of most-comprehensively connected banks. This core has become more important during the crisis, and the widened reserve target bands have allowed banks to exercise more discretion in forming relationships. However, when for a short time the core banks appeared risky, correspondents preferred to diversify and reduce their reliance on the core.

Chapter 11 (Galbiati and Soramäki) introduced agent-based modelling for a multi-agent, multi-period model of an RTGS payment system. Banks decide how much costly liquidity to allocate to the settlement process to execute an exogenous, random stream of payment orders. The analysis finds an equilibrium level of liquidity to be posted in the system, a liquidity demand curve which links liquidity to delay costs, and insights on the efficiency of alternative system configurations. For a wide range of costs, efficiency (measured by the netting ratio) could be enhanced if banks were to commit more liquidity than they do in equilibrium. This might constitute a rationale for imposing measures that encourage liquidity provision (eg throughput guidelines). Systems with fewer participants are found to be more liquidity-efficient than larger ones, due to the emergence of ‘liquidity pooling’ effects.

These studies show that the simulator research network has considerably expanded and diversified over time. This has increased our knowledge of payment systems and of the internal and external factors and parameters that affect them. The cooperation has also resulted in exchange of experiences and knowledge among central banks, in which the tools and methods are used in several countries and with an increased level of sophistication. This has led to the implementation of new and more robust risk mitigation facilities in payments systems, resulting in increased systemic stability. The Bank of Finland is grateful for the rewarding cooperation that underlies this publication series.

Due to the positive feedback, the Bank of Finland has decided to increase simulator support services in the form of consultancy, education, helpdesk and analysis services. The Bank of Finland will also continue the development of the simulation tool itself. Supporting the simulator research network via seminars and publications will also be a key objective. Network analysis studies, agent-based and other behavioural models, as well as market-linked payment flows are interesting emerging directions for research. The financial turmoil will probably also trigger a new brand of stability study that analyses payment system performance under exceptional financial conditions. It might be the case that the simulation tools will be employed and developed for more general stability analyses. We hope that this publication will stimulate new studies in this multidimensional business area and help to increase the efficiency and stability of payment systems.

Chapter 2

A quantitative assessment of international best practice for business continuity arrangements in payment systems

Kristian Sparre Andersen – Irene Madsen

2 A quantitative assessment of international best practice for business continuity arrangements in payment systems

Abstract

In recent years several central banks have studied how operational incidents affect settlement in payment systems. We expand previous studies by explicitly taking account of international best practices for business continuity arrangements. Specifically, we simulate operational incidents in Danmarks Nationalbank's RTGS payment system, Kronos, with the aim of identifying critical values vis-à-vis 1) recovery time, 2) critical participants, 3) contingency measures and 4) stop sending.

We find that oversight of payment systems can benefit from quantification of these critical values when assessing whether business continuity arrangements accord with international best practice. The simulations indicate that for Kronos resumption of normal settlement within 4 hours is sufficient and that a participant with a share of 1½–2 per cent of turnover should be classified as critical. Likewise, the simulations show that a capacity for settling 5–10 per cent of average daily transaction volume in contingency mode would significantly reduce the impact of operational incidents.

2.1 Introduction

2.1.1 Outline

In this section we introduce our study by presenting the questions we wish to address. In addition, we identify some aspects of business continuity where quantitative benchmarks for best practice can be established. This is followed by a brief survey in section 2.2 of previous research that assesses the impacts of operational incidents in payment systems.

Section 2.3 presents the key features of Danmarks Nationalbank's RTGS payment system, Kronos, and the types of transaction settled in the system. The section also gives an overview of incidents that have occurred in the system.

In sections 2.4 and 2.5 we present the results of the simulations, first for system-level incidents, then for participant-level incidents.

Section 2.6 concludes with the results of our simulations and especially interpretations of these in respect of international best practice for business continuity.

2.1.2 Motivation

Operational disruptions in payment systems can jeopardize financial stability. This is reflected in the core principles for systemically important payment systems issued by CPSS in January 2001.[1] In core principle no. 7, it is stated that payment systems 'should ensure a high degree of … operational reliability and should have contingency arrangements for timely completion of daily processing'.

These high-level objectives are elaborated in the core principles report which includes a number of more detailed (but still only qualitative) payment-system objectives aimed at securing a robust environment for processing payments. Specifically, business continuity arrangements should be in place that 1) minimise the risk of disruptions, 2) enable settlement of urgent payments in contingency mode during disruptions and, 3) shorten the period for restarting the system, if necessary, at a second site.

But how can central banks' oversight units be sure that a payment system's business continuity arrangements are adequate? Is it reasonable to base such assurance on quantitative measures? If so, what are the critical values for these measures? In the end, do these critical values accord with observed international best practice for business continuity arrangements?

With this in mind we will address the following questions:

- How fast should settlement resume after an operational incident occurs? (When do the risks and costs caused by operational disruptions become unacceptable)
- Who are the critical participants of the system? (Are there participants that can seriously disrupt settlement if hit by an incident)

[1] See Committee on Payment and Settlement Systems (2001).

- How many payments should the system operator be able to settle in contingency mode (eg by paper-based manual procedures before normal operations are resumed)?
- How fast should stop sending be applied on a participant unable to send payment orders? (ie to avoid that the participant absorb too much of the system's liquidity).

2.1.3 International best practice for business continuity

According to the CPSS report on Core Principles for systemically important payment systems, the methods and tools employed in the management of operational risk and business continuity in payment systems are still mainly qualitative in nature. Best practice is thus more a question of employing the specific methods and tools that will ensure the best possible operational risk management. Although quantitative methods and tools are still in their infancy as regards management of operational risks and business continuity, both in general and in payment systems, a number of quantitative benchmarks have evolved in the last decade.

Today, a payment system's disruption tolerance is typically defined in a service level agreement as the minimum acceptable availability during opening hours. Because central banks' RTGS systems are vital for the stability of the overall financial system the minimum acceptable availability are typically set above 99.5 per cent per annum.

In addition, the maximum time allowed for resumption of a system's normal operations should be based on the classification of (the severity of) operational disruptions in which the duration of the incident is an important (but not the only) factor. Eg, in Europe and USA systemically important payment systems should be able to recover and resume normal (or normal-like) operations within two hours after the occurrence of an incident.[2]

Since the terrorist attacks in New York on 11 September 2001, central banks have acknowledged that disruptions for certain participants can have severe impacts on the smooth functioning of a payment system. Hence central banks have begun to identify and define specific business continuity objectives for critical participants, eg by requiring such participants to be able to resume operations within two hours after an incident occurs.

[2] See ECB (2006) and Federal Reserve et al (2003).

Finally, the capacity of the payment operation department's staff for settling payments in contingency mode should be based on the level of urgent payments that must be settled in a timely manner so as not to have a major impact on financial stability or otherwise severely disrupt the financial system. (The authors are not aware of published quantitative requirements on the capacity of settling transactions in contingency mode within a specified timeframe while a payment system is unable to operate normally.)

2.2 Previous research

There are several studies that provide quantitative assessments of how operational incidents affect settlement in RTGS payment systems. Here we mention a few that in particular have motivated the research presented in this paper. However, none of the studies explicitly attempted to establish specific quantitative business continuity requirements for payment systems.

McAndrews and Potter[3] give a thorough presentation of the events following the terrorist attack on the World Trade Center on 11 September 2001. This tragedy made it possible to see how payment flows and liquidity holdings were affected for several days when several participants in a payment system in a financial centre were hit by a devastating incident. The attack caused a sharp drop in payments settled in the Federal Reserve RTGS system, Fedwire, on the day of the attack as well as a pronounced shift in the timing of settlements, due to delays in incoming payment orders, which led unexpectedly to participants' liquidity shortages. On the days after 11 September, there were also wide deviations from the normal settlement pattern in Fedwire because resumption of normal payment operations progressed slowly for a number of participants with Manhattan operations and because the clean-up required a substantial increase in payment volume for several days after the attack. In addition, McAndrews and Potter describe how, shortly after the attack, the Federal Reserve secured the financial system and thereby the payment flows by providing ample liquidity to participants that ran of out liquid funds.

In studies by Bedford et al[4] and Enge and Øverli,[5] operational incidents in the English CHAPS system and Norwegian NBO system

[3] See McAndrews and Potter (2002).
[4] See Bedford et al (2004).
[5] See Enge and Øverli (2006).

were studied via simulations with varying degrees of participants' liquidity holdings. Both studies found that participants' actual liquidity holdings had to be reduced significantly before settlement was severely affected by incidents at one of the large participants. In Bedford et al even with multiple simultaneous incidents at several participants the CHAPS system seemed to be quite robust as participants typically had ample liquidity holdings exceeding the system's upper bound of liquidity requirement.

Schmitz et al[6] studied the effect of operational incidents in the Austrian ARTIS system, in which both stop-sending rules and contingency measures were included in the simulations. The contingency measures envisaged included both transfers by the participant hit by an incident (using other means than the normal procedures, eg paper-based payment orders) and debit authorisations which allowed other participants to draw money from the account of the participant hit by the incident. Schmitz et al found that stop-sending improved the settlement more than contingency measures.

Lublóy and Tanai[7] analysed the resilience of the Hungarian payment system, VIBER, in much the same way as Schmitz et al (2006), by including both stop-sending rules and contingency measures in the simulations. In addition, they studied the effects of incidents which only caused disruptions for part of the system's opening hours.

As part of their research, Lublóy and Tanai made a very interesting comparison of VIBER with a number of other central banks' RTGS systems based on similar studies. This comparison shows interesting similarities and differences across payment systems in which two factors seem to dominate the resilience of payments systems against operational disruption: size of participants' liquidity holdings and the concentration of payments turnover in the systems.

[6] See Schmitz et al (2006).
[7] See Lublóy and Tanai (2008).

2.3 Payment settlement in Kronos

2.3.1 System features

Kronos is Danmarks Nationalbank's payment system, which has been operating since 19 November 2001.[8] The key features of the system are:

Real-time gross settlement (RTGS)
The system's real-time gross settlement mechanism enables immediate settlement of a payment after reception of the payment order, provided there are sufficient funds on the participant's RTGS account.

The system is open between 7:00 and 15:30 for interbank transactions, monetary policy operations and liquidity transfers to other systems; cf. section 3.2.

Liquidity transfers to the VP Settlement and the Sumclearing in preparation of the night settlement in these systems are executed between 16:00 and 16:30.

Transaction queue
A payment is placed in the systems' transaction queue at reception if funds on the participant's RTGS account are insufficient for immediate settlement. The standard release mechanism in the transaction queue is FIFO.

To handle situations where the first payment in the queue blocks subsequent (smaller) payments, participants can opt for a bypass function which allows settlement of subsequent payments if they do not exceed the available funds.

In addition, participants can manually, via the Kronos terminal, change the order of payments in the queue and cancel queued payments.

Payment orders in the transaction queue that are not settled before the system closes at 15:30 are deleted. Settlement will then only take place if participants retransmit the payment orders in question.

Gridlock resolution mechanism
Settlement can be effected in certain cases by activating the system's gridlock resolution mechanism. This mechanism can simultaneously

---

[8] See Angelius and Henneberg (2002).

release several participants' payments which are mutually awaiting each other's settlement, provided that no participants' accounts would subsequently be overdrawn. The gridlock resolution mechanism allows for settlement of an optimal subset of queued payments without violating the priorities of queued payments.

Due to the abundance of participants' settlement assets, it has so far not been necessary to resolve any gridlocks via this mechanism.

Standing orders
Participants can enter standing orders that are transferred daily from participants' RTGS accounts to settlement accounts for ancillary systems at preset times that match the settlement blocks in the ancillary systems.

Value date queue
Participants can enter payments into the system's value date queue for settlement up to 14 banking days later. On a payment's value date, it is automatically transferred to the transaction queue for settlement right after Kronos opens at 7:00.

2.3.2 Transaction types, settlement assets and participants

Kronos is used for large, time-critical payments in Danish kroner between banking institutions etc. In addition, the system is used for monetary-policy transactions, liquidity transfers to three ancillary systems – VP Settlement (securities), Sumclearing (retail payments) and the international currency-settlement system CLS.[9] Further, the system is used for cross-border transfers of collateral through the Scandinavian Cash Pool.[10]

Participants' settlement assets consist of funds on the participants' RTGS accounts in Danmarks Nationalbank and intraday credit provided by Danmarks Nationalbank against collateral. Intraday credits are applied as overdrafts on RTGS accounts. According to the terms and conditions for RTGS accounts, a participant is obliged to

[9] See Danmarks Nationalbank (2005).

[10] Scandinavian Cash Pool (SCP) is an automated system for pledging of cross-border collateral between Denmark, Norway and Sweden. The purpose of SCP is that liquidity raised in the central bank of one country can be pledged as collateral for intraday credit from the central bank of another country.

cover overdrafts so that there are positive funds on the RTGS account every banking day at 15:30.[11]

Participants in Kronos (holders of RTGS accounts in Danmarks Nationalbank) are Danish credit institutions and investment companies, including branches of foreign entities. Besides these institutions other entities which are important for the settlement of payments in Danish kroner can be allowed to open a RTGS account and participate in Kronos, eg CLS Bank. At the end of 2007 the system had 120 participants.

A detailed description of the data used in the simulations can be found in Appendix B.

Table 2.1 **Payment flows in Kronos in 2007**[12]

DKK billion	Total	Average per banking day	Per cent
Interbank transactions	30,876	124	44
Monetary policy operations	13,662	55	19
Miscellaneous	535	2	1
Liquidity transfers to other systems			
- VP Settlement	9,039	36	13
- Sumclearing	9,861	40	14
- CLS Settlement	3,460	14	5
- Scandinavian Cash Pool	2,827	11	4
Total	70,259	282	100

Source: Madsen (2008)

2.3.3 Frequency of operational incidents in Kronos

The relevance of simulating the impact of operational disruptions in a payment system depends on the perception of the risk of such incidents; cf. appendix A. An obvious way to form one's perception is to look at the actual occurrence of incidents in the system that has impacted the payment settlement process.

[11] See Danmarks Nationalbank (2005).

[12] All values in this chapter are expressed in Danish kroner (DKK). In January 2008 the DKK/USD and DKK/EUR average exchange rates were 506.24 and 745.05 (100 units) respectively.

Figure 2.1 **Incidents in Kronos (system level), 2005–2007**

Figure 2.1a Number of incidents in Kronos

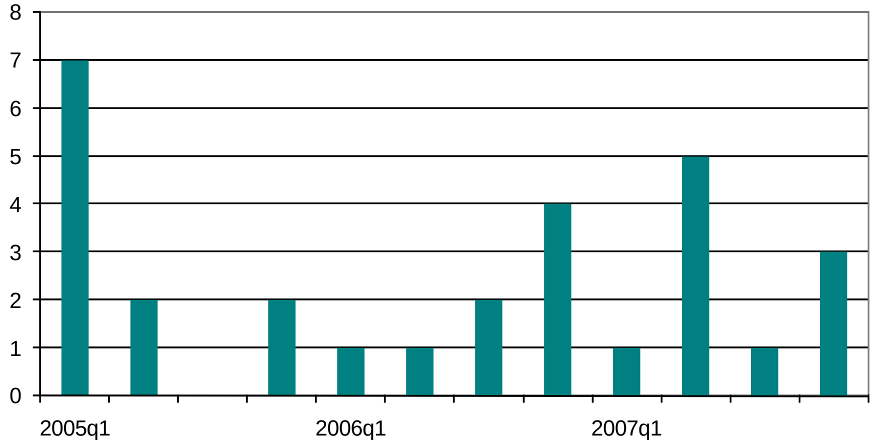


Figure 2.1b Average duration of incidents in Kronos

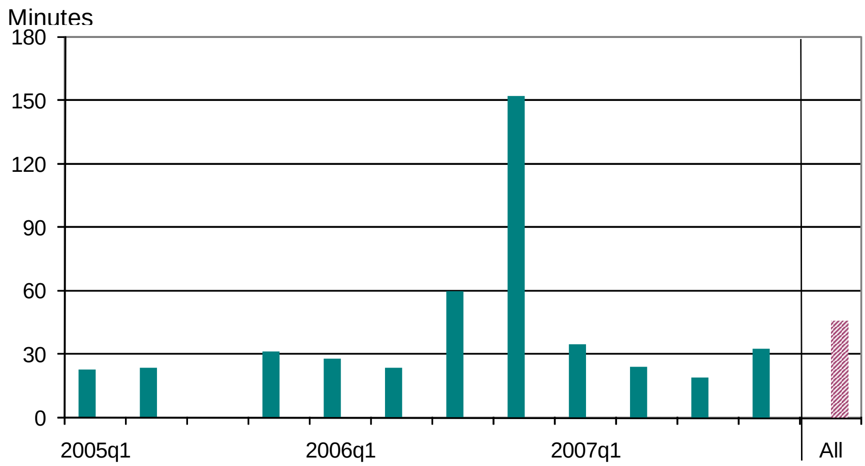


Figure 2.1c Start of incidents in Kronos

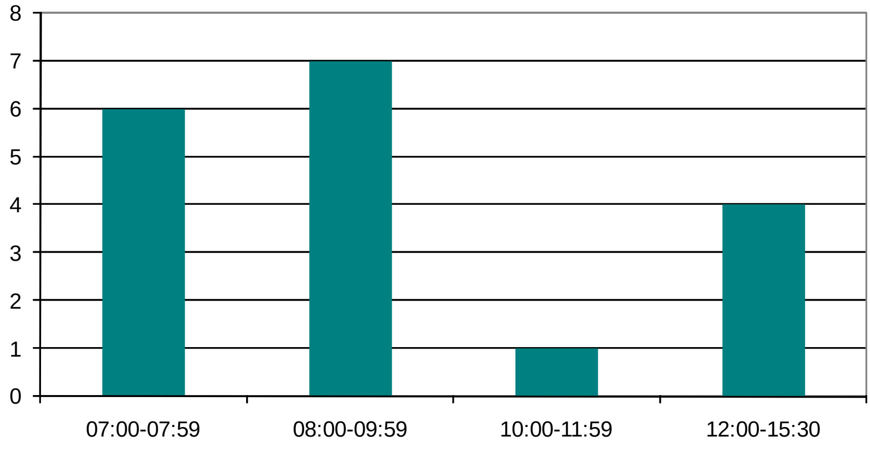
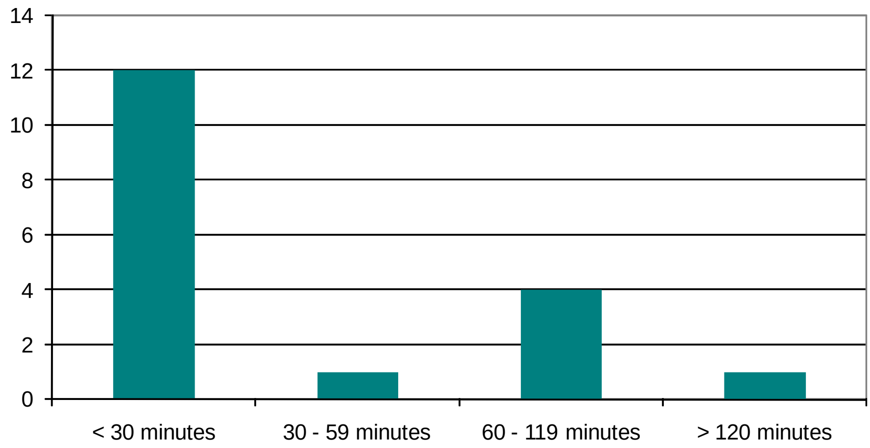


Figure 2.1d Duration of incidents in Kronos



Note: Figures 2.1.c and 2.1.d are based on incidents reported in 2006–2007.

In the years 2005–2007, Kronos experienced 29 incidents, or approximately 10 incidents per year. It should be noted that some minor incidents lasting a few minutes were not included in the statistics.

The average duration of the incidents recorded was around 45 minutes, with a notable peak in the second half of 2006. Excluding the incidents in the second half of 2006, the average duration of the remaining incidents was less than 30 minutes. Of the 5 incidents lasting more than 60 minutes in 2006–2007, 4 occurred in the second half of 2006. The root cause of the incidents in 2006 was problems with the IT supplier's connection to SWIFT; cf. appendix C.

The incidents occurred most often before noon. This happens to be the busiest time during Kronos' opening hours. Especially between 8:00 and 10:00, participants send many payment orders to the system for settlement (see Figure B.1).

Appendix C contains a description of two major incidents in recent years that in particular disrupted the settlement of payments in Kronos. The main impact of these two long lasting incidents was on the Danish money market. To normalise the market, Danmarks Nationalbank in both cases had to give participants extraordinary access to liquidity.

2.4 Simulation of incidents in Kronos – System level

Incidents at system level prevent settlement of payments completely if the system operator cannot provide participants with contingency measures and the participants are unable to settle payments by other means, eg in the correspondent banking network. Due to the size of payments settlement in central bank-owned RTGS systems like Kronos, the correspondent banking network has some disadvantages, as settlement then will take place in commercial bank money which, among other things, usually means higher settlement risk compared with settlement in central bank money.[13] For many participants, central bank provision of settlement has the additional advantage that settlement then does not take place at a competitor, ie settlement takes place on neutral ground. RTGS systems like Kronos are therefore obliged to implement effective contingency measures that can ensure quick resumption of settlement if the main system cannot operate.

The question then becomes how to decide what requirements effective contingency measures in payment systems should fulfil so as to ensure that incidents do not cause unnecessary operational and financial costs for participants and the central bank.

[13] See Committee on Payment and Settlement Systems (2003).

To start, we look at incidents that stop normal settlement in Kronos for a whole day, and then we move on to how incidents of shorter duration affect the settlement process.

2.4.1 Effect of contingency measures on whole-day incidents

The purpose of contingency measures is to limit risks and overall costs associated with operational incidents; cf. appendix A. Thus the scale of contingency measures the system operator relies on when the system is not operating normally depends not only on the number of payments settled on a daily basis but also on the size distribution of the transactions.

Figure 2.2 **Value of contingency payments, per cent of daily total payments**

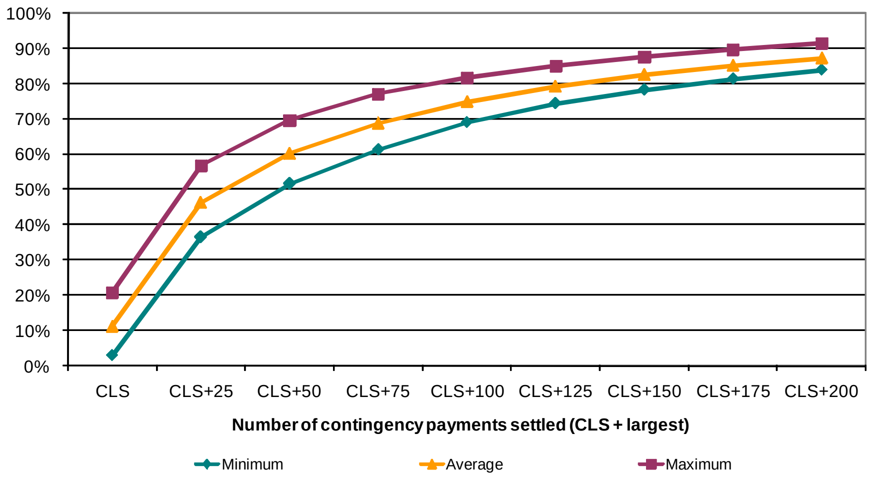


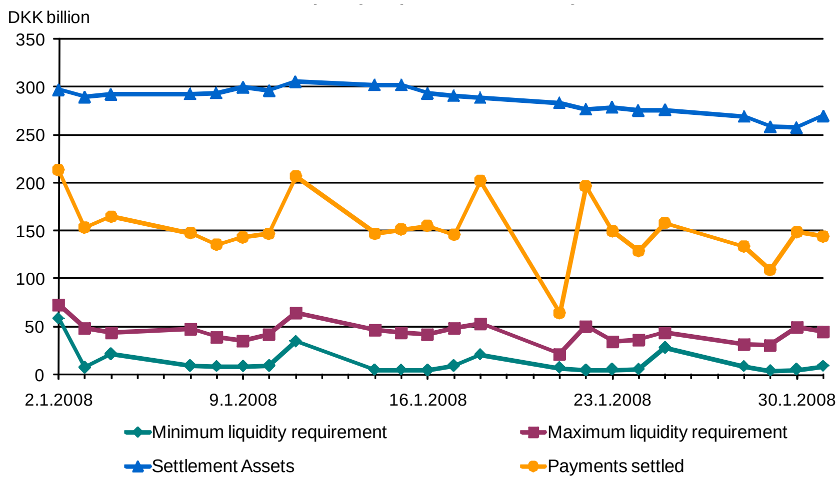
Figure 2.2 shows the capacity of potential contingency measure levels consisting of payments related to the CLS settlement (always very time-critical!) and a number of the largest payments in case of operational incidents.[14] On all days the settlement of CLS pay-ins/pay-outs and the 50 largest transactions would ensure settlement of 50 per cent or more of the value participants planned to settle that day. If the

[14] CLS and the 200 largest transactions amounted in January 2008 to approx. 10 per cent of the daily volume settled in Kronos.

contingency measures were expanded so that the operator could handle the 200 largest transactions, in addition to the CLS transactions, the measures would cover 84–91 per cent of the value participants planned to settle.

The remaining payments not settled due to a whole-day incident would not have, on any single day, exceeded DKK 27 billion (13 per cent of the value settled that day in Kronos). Although this is a substantial amount, one should bear in mind that commercial banks in Denmark had DKK 1.046 billion of liquid assets at end-2007.[15] Consequently, the banking system in general has ample holdings of liquidity to cover unanticipated stops in the incoming liquidity flows. In other words, payments not settled due to a whole-day incident would, with the most ambitious contingency measures we examined, amount to less than 3 per cent of the participants' liquid assets.

Figure 2.3 **Payments settled compared with participants' settlement assets and liquidity requirements, January 2008**



Note: Participants' overall daily minimum liquidity requirement equals the system's 'lower bound' (amount of liquidity that participants need to settle all payments at the end of the day; calculated by using a multilateral netting algorithm). Participants' overall daily maximum liquidity requirement is equal to the system's 'upper bound' (amount of liquidity that participants need to settle all payments without delay; calculated as the maximum negative difference between incoming and outgoing payments for each participant during Kronos' opening hours aggregated over all participants).

[15] See Finanstilsynet (2008).

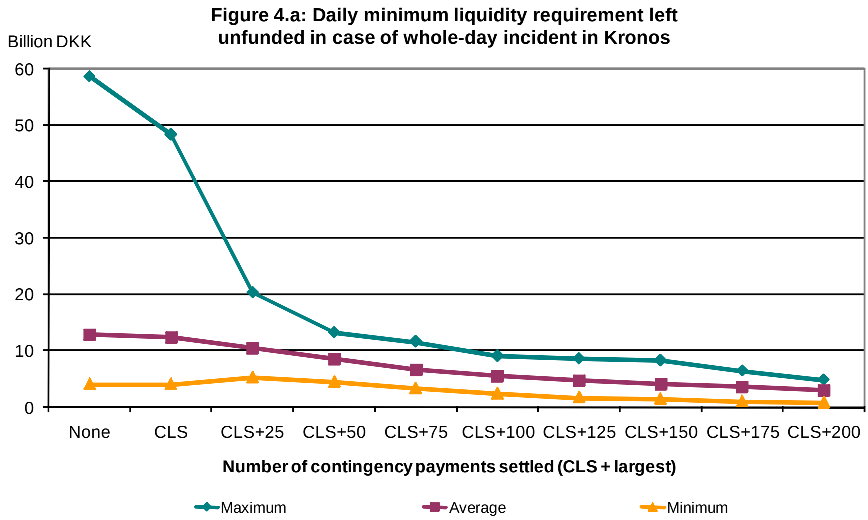
When setting the requirement for contingency measures, the system operator should look not only at how these perform in relation to the total value of payments settled in the system but also at how these mitigate the impact of incidents on participants' holdings of settlement assets and liquidity requirements. On average, the participants' holdings of settlement assets were nearly 2 times larger than the average daily values settled (DKK 286 and 152 billion, respectively). As a consequence, the participants have a substantial capacity to withstand delays in incoming payments, but only on an intraday basis, as the settlement assets in January 2008 consisted of 98 per cent of intraday credits granted against collateralised securities and only 2 per cent of funds in RTGS accounts (see table B.1). Participants' daily minimum liquidity requirements were on average DKK 13 billion (varying between 59 billion and 4 billion) while their daily maximum liquidity requirements on average amounted to DKK 44 billion (varying between 73 and 21 billion). With regard to the minimum liquidity requirement, it should be noted that four of the five peaks in figure 3 occurred on Fridays on which Danmarks Nationalbank conducted weekly market operations. The first and highest peak (in the figure) occurred on 2 January when the activity in Kronos was affected by a large refinancing of adjustable-rate mortgages bonds, instalments on housing loans and tax payments.

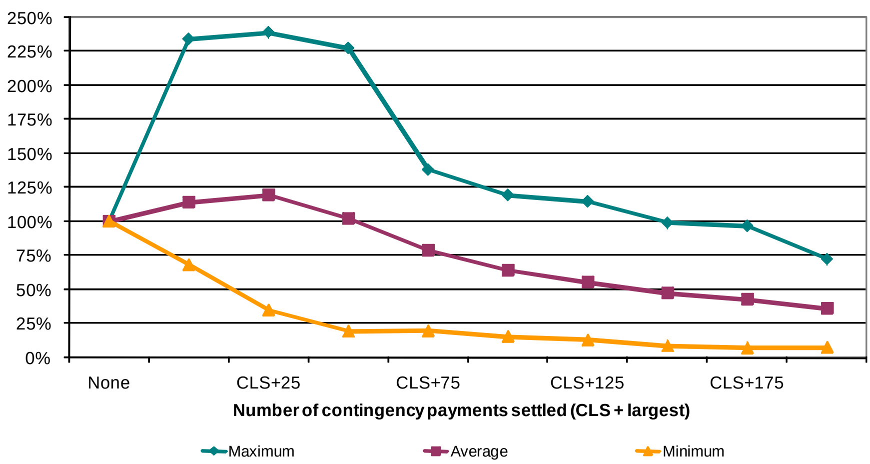
The contingency measures we examined covered on average 75 per cent of participants' liquidity requirements (see figure 4.a). The capacity of contingency measures to fund participants' liquidity requirements was especially good on days when the participants had large liquidity requirements, eg on 2 January 2008.[16]

A closer look at the effect of contingency measures in figure 4.b, however, reveals that low levels of contingency measures in some cases increase, rather than reduce, the participants' liquidity requirements. In the worst case, settling CLS pay-ins/CLS pay-outs and the 50 largest transactions in contingency mode more than doubled the participants' liquidity requirements. Further, on one day settlement of CLS payments plus the 200 largest payments would only cover 27 per cent of the minimum liquidity requirements needed for settling all planned payments. However, this effect only occurred on

[16] The net liquidity participants receive via payment settlement in Kronos is typically used for payment obligations in ancillary systems or invested in the money market (including certificates of deposit issued by Danmarks Nationalbank). Consequently, participants' funds in RTGS accounts when Kronos closes at 15:30 are quite stable over time despite the large intraday fluctuations.

Figure 2.4 **Daily minimum liquidity requirement left unfunded in case of whole-day incident in Kronos**





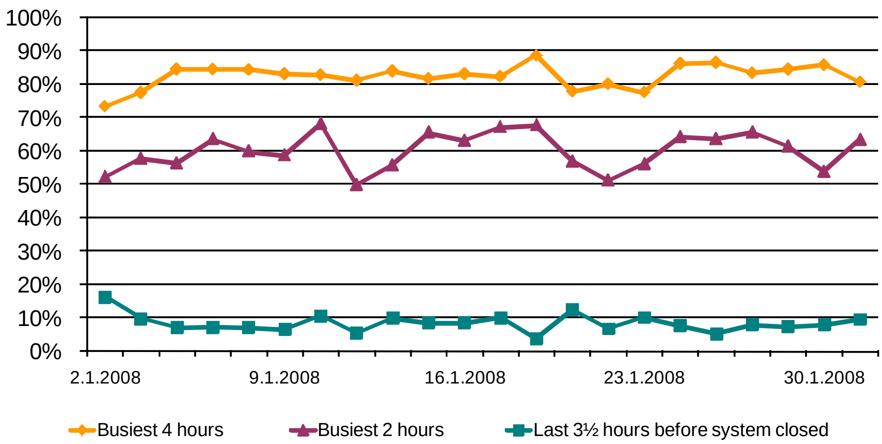
Note: The minimum liquidity requirement left unfunded in case of whole-day incident in Kronos equals the minimum liquidity requirement (lower bound) for payments not settled via contingency measures.

days when participants' minimum liquidity requirements were modest (less than DKK 8 billion) relative to the daily turnover in the Danish short-term money market.[17]

2.4.2 Incidents of shorter duration

We also studied the effects of three types of incidents disrupting all settlement in Kronos for less than one day. First, we analysed to what extent payments would be delayed if the system could not operate in the busiest two hours and four hours on each banking day in January 2008; these 'rush hours' in Kronos typically occur before noon, the busiest 2 hours being in the time frame 8:00–11:00 and busiest 4 hours in 7:00–12:00. Second, we analysed the impact of an incident that halted settlement at noon and where resumption of normal settlement could not occur before Kronos' closing time at 15:30.

Figure 2.5 **Settlement turnover during busiest opening hours and end-of-day, per cent of total turnover**



Incidents that halt settlement in the busiest two hours would affect on average 60 per cent of the day's transactions. On the day with

[17] According to statistics compiled by Danmarks Nationalbank's Market Operations Department, daily turnover in the Danish short-term money market was DKK 51 billion in January 2008.For a description of the Danish money market see Abildgren and Arnt (2004).

maximum impact of a 2 hour disruption, transactions of DKK 137 billion would be affected, ie 68 per cent of settled value. Similarly, incidents that halt settlement in the busiest 4 hours would affect on average 82 per cent of the day's transactions. On the day with the maximum impact of a 4 hour disruption, transactions of DKK 179 would be affected, ie 86 per cent of settled value.

The presence of highly urgent payments, eg CLS pay-ins/pay-outs, implies that contingency measures have to be in place to ensure quick settlement of a number of transactions. However, assuming that normal settlement resumes several hours before Kronos' closing time, it seems that a 4 hour disruption of settlement will not be excessively harmful. This is due to the fact that the system, when functioning normally, has the capacity to process and settle all transactions, even on the busiest day, within 1 hour. This was confirmed by the incident on 16 November 2006 described in appendix C. It is also Danmarks Nationalbank's impression from this and other incidents that the number of highly urgent payments is relatively modest.

Participants' decision to forward payment orders to Danmarks Nationalbank when Kronos is not operating normally also depend on the costs associated with contingency measures. Based on participants' behaviour when the system was not operating normally due to incidents, it is clear that they are reluctant to (manually) remove payment orders from systems that automatically send payment orders to Kronos for execution and send the payment orders to Danmarks Nationalbank by fax (also manually). The participants' automated systems typically generate payment orders that include information to enable automatic handling of payment orders when received by other participants. An essential prerequisite for this 'Straight-through-processing' of transactions in Kronos is that the vast majority of payment orders are based on standard SWIFT messages (MT103 and MT202).[18,19]

As regards incidents starting at noon and lasting the rest of the day, 8 per cent of the transactions value would have been affected on average. On the day with the maximum impact of an end-of-day incident, transactions of DKK 35 billion would have been affected, ie 16 per cent of total value settled on the day in question.

[18] See Danmarks Nationalbank (2005).

[19] One should bear in mind that these cost considerations do not include interest costs if 1) the participants' settlement assets at Danmarks Nationalbank are not exhausted and 2) normal operations are assumed to resume before end-of-day, as there is no interest charge on intraday credits provided by Danmarks Nationalbank.

Based on these facts, when considering the settlement activity in Kronos, it does not seem crucial that full-scale contingency measures are in place soon after an incident halts settlement. A time-frame of 4 hours for resuming full-scale contingency measures (normal-like operations) seems sufficient for Kronos, as manual settlement of less than 250 transactions while the system is not operating normally would account for around 80–90 per cent of any day's total payments value (see figure 2.2).

Participants' emphasis on maintaining the straight-through processing of payment orders as described above favours a quicker resumption of normal settlement than 4 hours. However, it can be difficult to justify investments in systems that can guarantee quick resumption of normal settlement, say within 2 hours, to save the costs associated with the settlement of up to 250 payments a few times during the system's lifetime.

In any case, it is crucial that manual-based contingency measures are applied as quickly as possible because of the presence of a number of urgent time-critical payments every day, eg payments to and from CLS Bank that cannot wait for start-up of full-scale contingency measures or resumption of normal operations.

2.5 Simulation of incidents in Kronos – Participant level

After having studied the impact of operational incidents on system level we now turn to incidents at participant level. In a historical perspective the root causes of many operational disruptions in payment systems lie with the participants, eg the terrorist attack on the World Trade Center on 11 September 2001.[20]

For quantifying how operational incidents at participants affect the settlement in Kronos we selected 10 of the most active participants in the system in terms of turnover and number of connections to other participants. All of these participants are active in the Danish money market. Four of the participants for which we simulated incidents are also liquidity providers in Danish kroner in the CLS settlement of FX contracts, of which 3 belong to foreign banking groups.

In this section we restrict the simulations to operational incidents lasting a whole day. In the simulations we study the effect of two

[20] See McAndrews and Potter (2002).

types of reactions to mitigate the impact of a large participant unable to send payment orders for settlement in Kronos due to an operational incident: 1) settlement of out-going payments for the affected participant via manual-based contingency measures and 2) the other participants securing their holdings of settlement assets by stopping the sending of payments to the affected participant.

In interpreting the simulation results below one should bear in mind that we do not take account of more complex patterns of reactions that take place outside the payment system, eg other participants choosing not to enter into financial transactions involving settlement of payments in Kronos vis-à-vis the affected participant on that particular day.

2.5.1 Critical participants

Settlement in Kronos is dominated by a few participants that are clearly critical for the smooth functioning of the system, as shown in table 2.2. However, besides the share of turnover, the impact of an incident is of course also highly dependent on the other participant's holdings of settlement assets.[21]

Moreover, incidents at participants with direct links to CLS typically have the greatest impacts of all in the simulations, not only because of their share of turnover, but also because CLS by nature only posses the funds it receives during a CLS settlement cycle and never pays out from its account in such a way that it would go into deficit. This is illustrated by Bank 4, which on one day in January 2008 would have had an exceptionally large impact on the system compared with its share of the turnover. Because a large pay-in from Bank 4 to CLS was not settled, the simulated incident created a larger second order impact than Bank 2 had on any day when removal of its transactions was simulated.

[21] See Bedford et al (2004).

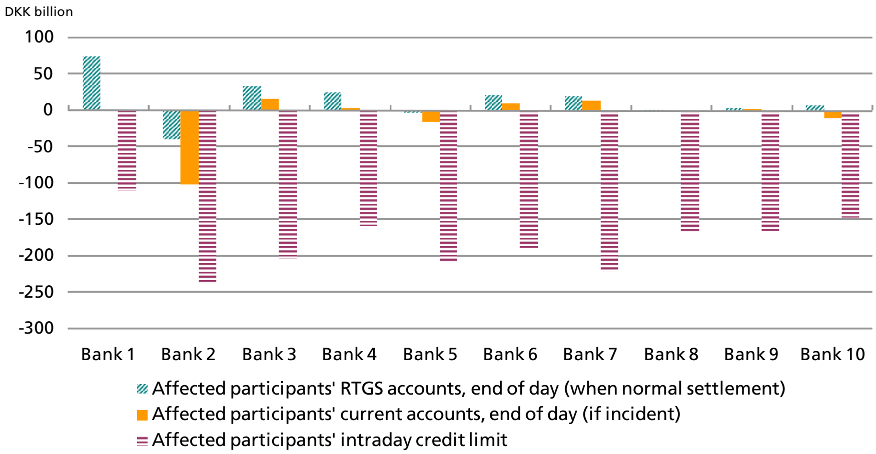
Table 2.2 **Worst case impact on total settlement in case of whole-day incidents at selected participants**

DKK billion Participant hit by incident	Normal settlement (no incident)	Direct impact	2nd order impact	Total impact	Payments settled	Payments settled, Per cent of normal settlement
Bank 1	207.1	72.8	13.7	86.5	120.6	58%
Bank 2	202.4	62.7	7.5	70.2	132.1	65%
Bank 3	207.1	16.7	3.8	20.5	186.6	90%
Bank 4	158.0	21.0	8.0	29.0	129.0	82%
Bank 5	197.0	12.7	0.1	12.8	184.3	94%
Bank 6	164.9	12.2	0.3	12.5	152.4	92%
Bank 7	213.5	3.8	2.3	6.1	207.3	97%
Bank 8	153.3	2.0	2.8	4.8	148.6	97%
Bank 9	147.5	1.2	-	1.2	146.3	99%
Bank 10	213.5	17.5	0.7	18.2	195.3	91%

Note: Direct impact is the value of payments not sent by participant affected by an incident; second order impact equals payments not settled for other banks due to insufficient funds on their RTGS accounts.

The impact on other participants' available funds on RTGS accounts at Danmarks Nationalbank for whole-day incidents at selected participants are shown in figure 2.6. The figure clearly illustrates that many of the affected participants still had a substantial cushion for meeting their payment obligations despite the fact that incidents in worst cases halted more than 40 per cent of the payments flow in the system. Of course, the existence of second order impacts means that a number of participants saw their holdings of settlement assets exhausted. (In all simulations we assumed that participants opted for the bypass function in the transaction queue).

Figure 2.6 **Impact on RTGS account end-of-day balances of whole-day (worst case) incident at participants**

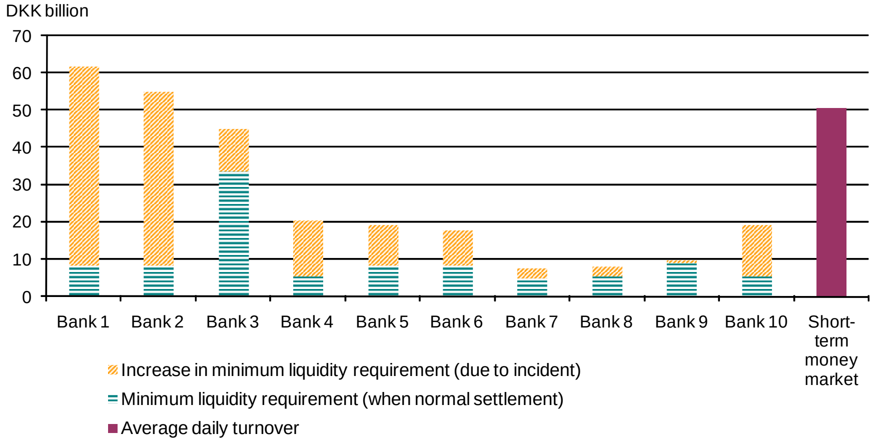


Note: The worst case incident for a participant is defined as the day when the participant created the largest decrease in RTGS account balances for the other participants compared with normal settlement.

As regards participants' liquidity requirements, almost all simulations showed that incidents at one of the two largest participants would leave the remaining participants with a substantial increase in their minimum liquidity requirement (see figure 2.7). On two days, an incident at Bank 1 or Bank 2 would even have entailed a larger minimum liquidity requirement for the remaining banks than on 2 January, which was the day in 2008 with by far the largest minimum liquidity requirement in Kronos (see figure 2.3).

For comparison, the average daily turnover in the Danish short-term money market is shown in figure 2.7, too. This reveals that incidents at the two largest participants are highly likely to cause extreme stress in the Danish money market. But incidents at 5 other participants seem to be capable of causing some stress in the money market in the absence of contingency measures or stop-sending rules. In interpreting figure 2.7 one should keep in mind that all participants depicted are active in the Danish money market. Incidents at one of these participants would therefore not only increase other participants' liquidity requirements but would also remove a player from an already fairly concentrated market.

Figure 2.7 **Increase in minimum liquidity requirement due to whole-day (worst case) incident at participants**



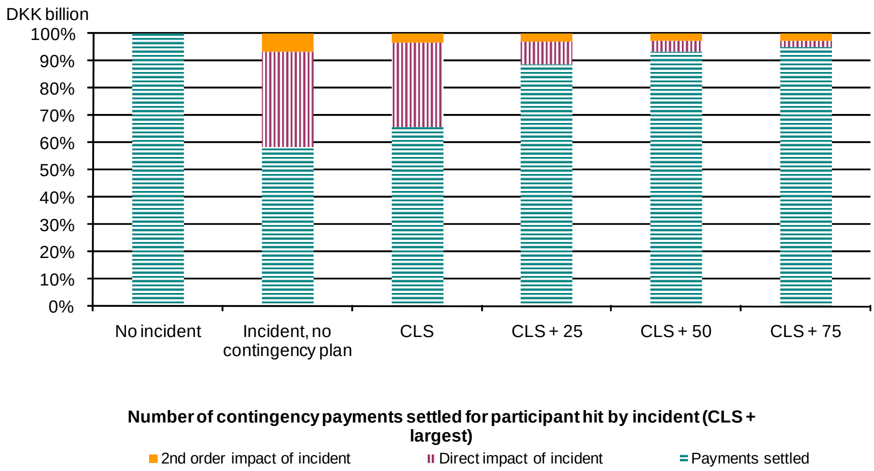
Note: The minimum liquidity requirement in case of an incident is the system's lower bound excluding transactions to and from the selected participant. The worst case incident for a participant is defined as the day on which the participant creates the largest increase in the minimum liquidity requirement for the other participants.

Assuming no contingency measures or stop sending-rules, there would be perhaps 7–8 candidates for critical participants in Kronos. Two participants, Bank 1 and Bank 2 (each accounting for over 20 per cent of system turnover), are each clearly critical, as they severely disrupt the settlement, in terms of the impact on total settlement, balances in RTGS accounts and minimum liquidity requirements. Looking at the worst case incidents, it seems reasonable to classify also Banks 3–6 and Bank 10 (with 3–6 per cent of turnover) as critical participants, both in terms of total impact on settlement and on how they can affect other participants' liquidity requirements if they cannot send payments to Kronos. When it comes to banks 7–8 (with 1½–2 per cent of turnover) it is surprising that an operational incident lasting a whole day in worst cases would entail second order impacts that are of the same magnitude as the direct impacts (see table 2.2). This is explained by the fact that these two banks have relatively larger payment flows with the smaller participants in the system that on average have less settlement assets available at Danmarks Nationalbank.

2.5.2 Effect of contingency measures

As already shown for incidents at system level, even rather modest contingency measures can notably reduce the impacts of incidents. As regards an incident hitting a participant, we simulated contingency measures including all CLS payments and the 25–75 largest payments. These constitute about a half of the number of payments in the previous section. (We assume that when Kronos technically functions normally, albeit with much lower activity, the system operator staff has to operate the settlement system as usual and, in addition, put resources into a contingency plan.)

Figure 2.8 **Effect of contingency measures on total settlement (worst case incident)**

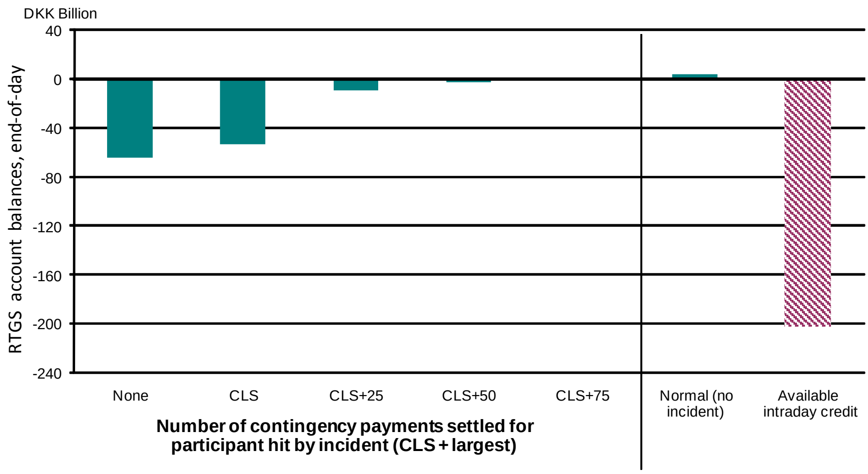


Note: Direct and second order impact is calculated as in table 2.2 by including settlement of a number of out-going payments in contingency mode for the participant hit by the incident.

Contingency measures' effect on total settlement when the largest participant is hit by an whole-day incident are depicted in figure 2.8. Settlement of CLS and the 25 largest transactions for the largest participant would have significantly dampened the effect of the worst case incident, as the impact would have been reduced with nearly 75 per cent compared with a situation where no contingency payments were settled.[22] This is also the case when looking on how contingency measures mitigate the impact on a) the balances of the other participants' RTGS accounts in figure 2.9 and b) the increase in their minimum liquidity requirements in figure 2.10.

On the basis of the simulations, it seems sufficient that the operator of Kronos can settle 50–75 payments in addition to all CLS pay-ins in contingency mode for a large participant unable to send payment orders to the system, as this will substantially reduce the impact of an incident.

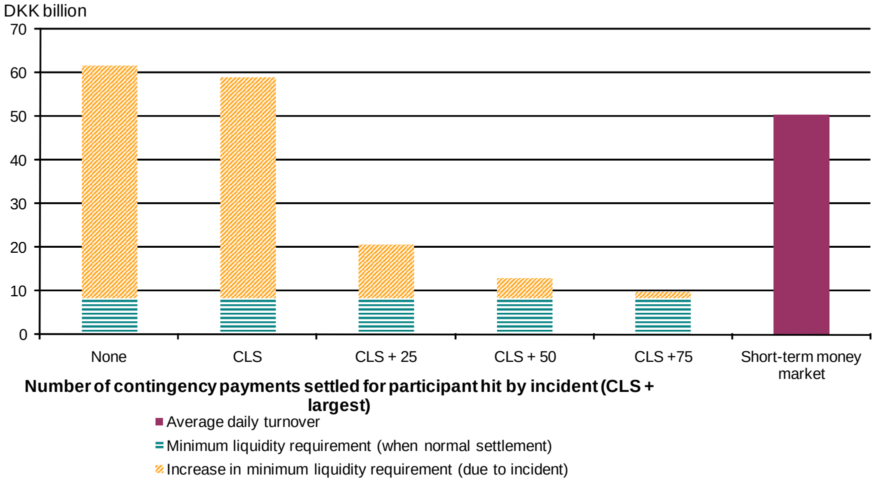
Figure 2.9 **Contingency measures effect on other participants' RTGS account end-of-day balances (worst case incident)**



Note: Effect of contingency measures on other participants' RTGS accounts are calculated as in figure 2.6 by including a number of transactions settled in contingency mode.

[22] It should be noted that in some simulations settlement of CLS transactions in contingency mode miti-gate the impact of the incident more than is the case in figure 2.8. Overall, however, settlement in contingency mode of the 25 largest transactions mitigated the impact of incidents on Kronos much more than settlement of the participants' CLS transactions.

Figure 2.10 **Effect of contingency measures on minimum liquidity requirement (worst case incident)**



Note: Minimum liquidity requirement is calculated as the system's lower bound excluding transactions to and from the participant hit by incident, except for transactions settled in contingency mode.

2.5.3 Effect of stop-sending rules

Another way to secure settlement in an RTGS system when a participant is hit by an incident is to activate a stop-sending rule which ensures that the participant in question does not drain the liquidity of other participants and thereby the system as a whole. Stop-sending is normally activated by the participant hit by the incident, either by sending a message to the other participants or to the system operator. The system operator can then decide or be obliged to temporarily close the account of the participant in question.

For the time being, the rules for participating in Kronos do not include a formal rule for stop-sending in the case of an operational disruption at one of the participants.

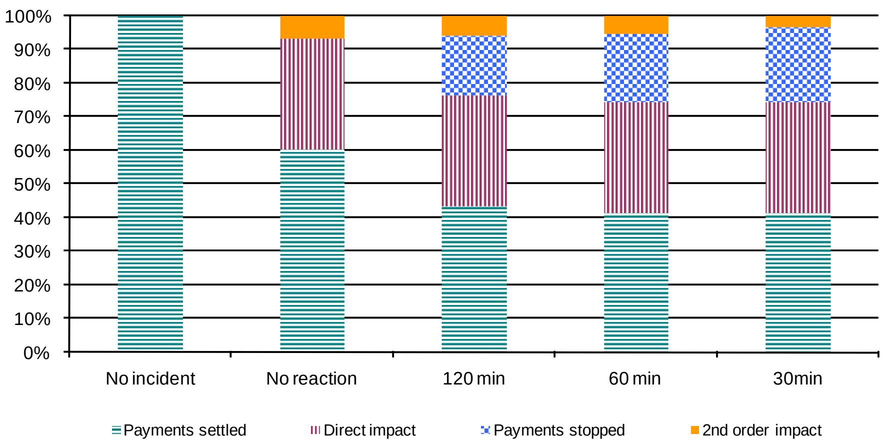
Stop-sending and contingency measures affect settlement quite differently. Whereas the objective of contingency measures discussed above is to keep settlement in the system as close to the original payment schedules as possible, including payments to and from the participant hit by an incident, stop-sending aims at excluding altogether this participant from the settlement. This is shown in figure 2.11 where the number if payments settled is the smaller, the sooner (more effectively) the stop-sending is activated. The same effect is seen in figure 2.12 where the depletion of funds on other participants' RTGS accounts is the less, the faster they react and stop sending payments to the participant hit by the incident.

The advantage of stop-sending as compared to contingency measures is that it is easy to apply and manage for the system operator.

But if the operational disruption should last for a longer time, stop-sending loses its advantage and the participant in question must somehow be involved in the settlement again. If this happens too late in the day, the backlog of deferred payments can be so large that it complicates the clearing up after the incident.

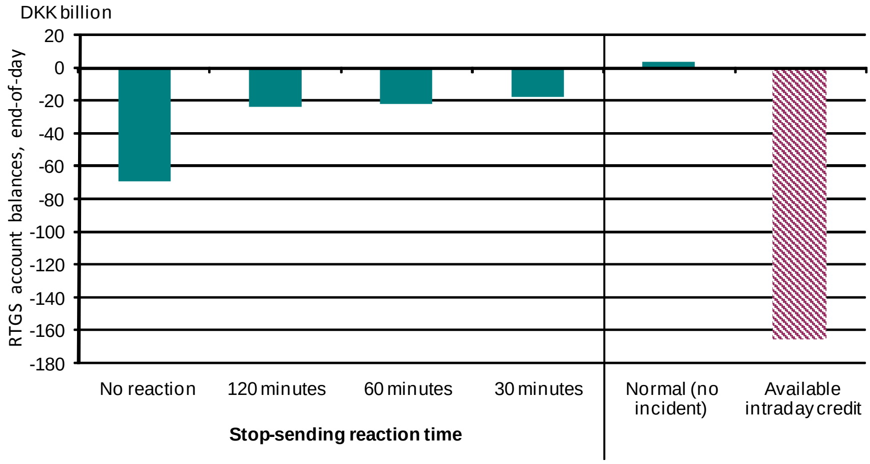
With this in mind, it seems that stop-sending, for the sake of business continuity, should only be applied for a shorter period, after which the participant hit by the incident and the system operator should switch to contingency measures of some kind, as envisaged in the simulations above.

Figure 2.11 **Effect of stop-sending reaction times on total settlement (worst case incident)**



Note: Direct and second order impacts are calculated as in table 2.2 by taking account of the fact that other participants stop sending payments to the participant hit by an incident after a specified reaction time.

Figure 2.12 **Effect of stop-sending reaction times on other participants' RTGS accounts (worst case incident)**



Note: Balances on other participants' RTGS accounts are calculated as in figure 2.6 by taking account of the fact that they stop sending payments to the participant hit by an incident after a specified reaction time.

2.6 Conclusion

Simulations of operational disruptions in the settlement of payments in Danmarks Nationalbank's RTGS system, Kronos, give us important insight into how to apply international best practice for business continuity arrangements.

The existence of a number of very urgent payment transactions on almost every banking day – eg CLS pay-ins and pay-outs – implies that Danmarks Nationalbank and critical participants should be able to switch to settlement via contingency measures very quickly. A switch to large-scale contingency measures that enables normal-like settlement, however, need only be completed within 4 hours to ensure that the backlog of payments does not become excessively large and pose a threat to the orderly ending of operations in Kronos within normal opening hours.

In case of an incident either at system level or at a large participant, settlement of all CLS transactions and the 50 largest transactions would amount to a substantial share of daily settled values. But this level of contingency measures would, in the case of an incident at system level, not be sufficient, as it sometimes increases, rather than reduces, the liquidity requirements participants would not get funded via payment settlement in Kronos.

If the capacity for settlement in contingency mode is expanded by an additional 150 large transactions, 84–91 per cent of payments turnover value would be settled. In this case, the negative impact of operational disruptions on participants' available funds in RTGS accounts and their ability to fund liquidity requirements would also be reduced to a level that seems manageable compared with the size of the Danish short-term money market. Therefore, settlement of all CLS transactions and the 200 largest transactions seems sufficient to minimise the risk of operational disruptions in Kronos which can jeopardise the stability of Denmark's financial system.

The two largest participants in Kronos are clearly critical participants, as they would inevitably have a huge impact on the settlement if they are not able to send payment orders to the system. A number of other participants with 3–6 per cent shares of payments turnover would on some days, if disconnected from the system, also have a significant impact on the payment flows. Consequently, they too should be classified as critical for the smooth functioning of the system. Further, two participants each with 1½–2 per cent share of turnover might need to be classified as critical, not because of the total impact if they are disconnected from the system, but because their position in the payment network surrounding Kronos gives rise to a larger second order impact than many of the larger participants. The large second order impacts of incidents at these two participants must be explained by the fact that they are more interconnected with other participants that have small holdings of settlement assets relative to their payments turnover compared with some of the other participants for which we simulated incidents at participant level.

As regards stop-sending rules, the simulations show that they can be quite effective for ensuring that a participant hit by an operational incident does not drain the system's liquidity. However, stop-sending rules cannot stand alone and should always be supplemented by contingency measures that are activated if the participant is not able to resume operations before Kronos closes down at the end of day.

Before one can conclude as to quantitative requirements for the business continuity arrangements in relation to Kronos, further simulations would be needed. These simulations should in particular look on incidents lasting more than one day, eg 3–5 banking days. In addition, our research would benefit from stress testing Kronos' robustness by reducing the amount of settlement assets available at the participants and by simulating incidents affecting several large participants. Eg, the relatively modest second order impacts shown in section 2.5.1 can be largely explained by the fact that the participants' settlement assets in January 2008 amounted to almost twice the average daily value of payments settled in the system.

References

Abildgren, K – Arnt, H (2004) **The Activity in the Danish Money Market.** Danmarks Nationalbank, Monetary Review, 2. Quarter.

Angelius, T – Henneberg, A (2002) **Danmarks Nationalbank's New Payment System, Kronos.** Danmarks Nationalbank, Monetary Review, 1st Quarter.

Basel Committee on Banking Supervision (2003) **Sound Practices for the Management and Supervision of Operational Risk.** Bank for International Settlements.

Bedford, P – Millard, S – Yang, J (2004) **Assessing operational risk in CHAPS Sterling; A simulation approach.** Financial Stability Review, June 2004, Bank of England.

Board of Governors et al (2003) **Interagency Paper on sound practices to strengthen the resilience of the US financial system.** Issued by Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency and Securities and Exchange Commission, April 7, 2003.

Committee on Payment and Settlement Systems (2001) **Core principles for Systemically Important Payment Systems.** Bank for International Settlements.

Committee on Payment and Settlement Systems (2003) **The role of central bank money in payment systems.** Bank for International Settlements.

Danmarks Nationalbank (2004) **Financial Management at Danmarks Nationalbank.**

Danmarks Nationalbank (2005) **Payment Systems in Denmark.**

Danmarks Nationalbank (2006) **Annual Report and Accounts 2006.**

Danmarks Nationalbank (2008) **Financial stability 2008.**

ECB (2006) **Business Continuity Oversight Expectations for Systemically Important Payment Systems.** European Central Bank.

Enge and Øverli (2006) **Intraday liquidity and the settlement of large-value payments: A simulation-based analysis.** Economic Bulletin 1/2006, Norges Bank.

Finanstilsynet (2008) **Statistical Material for Danish Banks 2007.** Danish Financial Supervisory Authority, www.ftnet.dk (in Danish only).

Jorion, P (2000) **Value at Risk: The new benchmark for managing financial risk.** 2nd Edition, McGraw-Hill.

Madsen, I (2008) **The Financial Sector's Payments via Kronos.** Danmarks Nationalbank, Monetary review, 1st Quarter.

McAndrews, J – Potter, S (2002) **Liquidity effects of the events of September 11, 2001.** Economic Policy R, Vol. 8, No. 2, Federal Reserve Bank of New York.

Lublóy – Tanai (2008) **Operational disruptions and the Hungarian Real Time Gross Settlement System (VIBER).** Occasional Papers 75, Magyar Nemzeti Bank.

Rørdam, K B – Bech, M L (2008) **The topology of Danish Interbank money flows.** Working paper 59, Danmarks Nationalbank.

Schmitz et al (2006) **Operational risk and contagion in the Austrian large-value payment system ARTIS.** Financial Stability Report 11, ÖNB.

Appendix A

Definition, measurement and management of operational risk in payment systems

Among the most important risk factors in payment systems are the operational risk factors. When one assesses operational risk in payment systems the methods and tools applied are often qualitative. Quantitative methods can, however, provide figures on certain dimensions of the operational risks of payment systems, which can be valuable inputs for system owners and overseers in the overall risk assessment.

Operational risks can be defined as the risks of economic loss resulting from inadequate or failed internal processes and systems, human errors, or from external events such as natural disasters, terrorism, etc.[23]

Economic losses due to operational incidents can take various forms depending on how the incidents affect the system and its participants:

- When an operational incident in a payment system results in loss of hardware or software, these will typically take the form of repair and investment costs of reconstructing or improving the part of the system hit by the incident. Repairs and investments may be required both in the system itself and at participants.
- Incidents can cause further costs at participants when they a) try to settle payments by other, less-efficient means, ie manual-based contingency measures; b) cancel or do not enter into profitable financial transactions because the payments associated with the transactions cannot be settled; or c) are liable for penalties as payments are not settled according to contractual obligations.
- In addition, incidents can due to delays and cancellations of payments result in unexpected credit and liquidity exposures at participants, which, depending on the circumstances in the financial markets, can lead to losses. This could be the case if participants encounter liquidity shortages and have to obtain liquidity from other sources on short notice, which can entail increased funding costs. Worst-case incidents may generate credit costs if one of the participants defaults while it has out-going payments in queue for settlement.

[23] See Basel Committee on Banking Supervision (2003).

- Finally, incidents can lower the demand for a payment system's services and thereby undermine the system's business case. This may happen if incidents develop into a reputational risk.

Inspired by methodologies applied when quantifying credit risk and market risk, eg Value-at-Risk, operational risk can be assessed as a function of the probability of incidents occurring and the impact of the incidents when they occur.[24] The impact of an incident could be one or more of the types of the economic losses mentioned above.

A key element in Core Principles 7 is that measures should be implemented and maintained with the aim of minimizing the operational risk in payment systems. Without taking account of these measures the result of the assessment would be the gross operational risk; where including the measures the assessment leads to the net operational risk, often termed the residual risk.[25]

Due to the complexity of the design of payment systems and the large number of threats to the security and operational reliability of the systems, a wide variety of measures have to be implemented. The purpose of these measures should be to reduce the probability of incidents as well as the impact of realised incidents. Reduction of operational risk can take the form of either measures that make the system less prone to disruptions or contingency measures that can be switched to if an incident in any way disrupts settlement in the main system.

The impact of an incident will depend on where the incident happens. An incident hitting the core system and thereby disrupting all settlement will of course have a greater impact than if the incident only hits one participant. Ie, the size of the initial incident also matters in assessing the operational risk caused by the incident. However, as already noted, settlements in most payment systems are dominated by a few critical participants. This explains why oversight of payment systems in recent years has expanded and now also encompasses the critical participants of the systems.

Timing and duration of incidents also matters in assessing the operational risk in payment systems, as payment settlement typically fluctuates during the systems' opening hours.

[24] See Jorion (2000).
[25] See Danmarks Nationalbank (2004).

Appendix B

Data used in simulations

Simulations were based on payment transactions in Kronos settled during 22 banking days in January 2008 (see table 2.1). The data sets used in the simulations consist of:

- Time-critical payments (RTGS transactions) between participants during Kronos' opening hours (7:00–15:30).
- CLS pay-ins and pay-outs.
- Funds on participants' RTGS accounts at start-of-day.
- Intraday credit limits granted to participants based on value of assets in securities accounts in VP Securities Services, the Danish central securities depository, and pledged to Danmarks Nationalbank.

Table B.1 **Overview of Payment transactions in Kronos, January 2008**

	Volume	DKK billion
Total turnover	63.467	3.348
– of which CLS pay-ins and pay-outs	2.144	391
Daily turnover		
– Average	2.885	152
– Minimum	2.605	64
– Maximum	3.905	213
Settlement assets (average start-of-day)		286
– of which balances on RTGS accounts		6
– of which intraday credit limits		280
Participants (active in January 2008)	117	

In table B.1 it should be noted that participants' average holdings of settlement assets exceeded the payment turnover on every day in January 2008. This has of course a stabilising effect on the simulation results.

As in many other RTGS payment systems operated by central banks in small European countries, the network topology of Kronos shows a pronounced concentration on participants active in FX settlement and the money market (see table B.2). 86 per cent of payments in Kronos involve one of the CLS liquidity providers while only 2 per cent involve neither CLS liquidity providers nor active money market participants.[26]

Table B.2 **Network topology of Kronos, January 2008 (simplified)**

Share of turnover in per cent	Receiving participants				
Sending participants	CLS	CLS-LP	MM-P	Other	All
CLS		6			6
5 CLS liquidity providers (CLS-LP)	6	48	10	3	67
8 money market participants (MM-P)		10	2	5	17
103 other participants		3	5	2	10
All participants		68	17	10	100

We excluded payments related to Danish ancillary systems, as virtually all of these are settled at night and hence seldom significantly alter participants' liquidity positions (funds available for payment transactions) during Kronos' normal opening hours. Participants' RTGS transactions in Kronos are of course closely related to the net payments settled in the ancillary systems during the night as part of participants' over-all liquidity management.

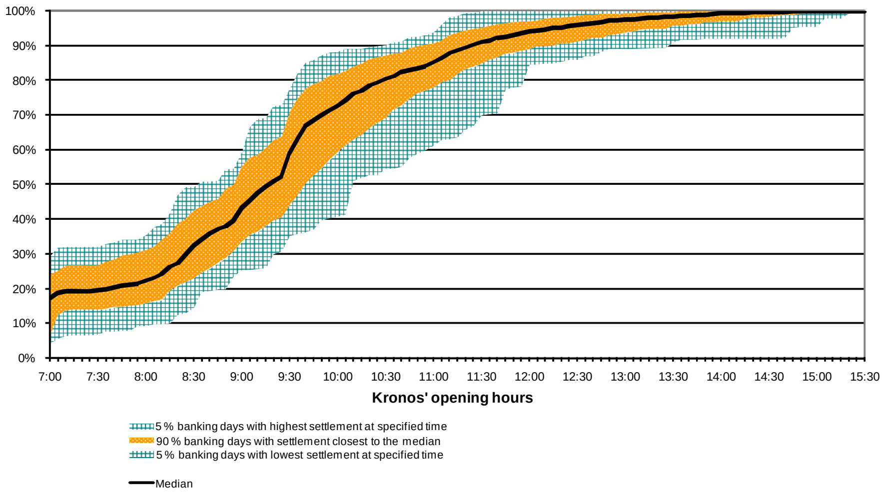
Similarly, monetary-policy operations were excluded, since these fully collateralised operations do not alter participants' holdings of available settlement assets. Assets eligible for loans in Danmarks Nationalbank and certificates of deposit issued by Danmarks Nationalbank are also eligible for obtaining intraday-credit. Consequently, monetary-policy operations merely move funds between participants' RTGS accounts and their holdings of monetary-policy instruments but leave their aggregate position at Danmarks Nationalbank unchanged.

Finally, the value of assets in the collateral accounts that participants use for obtaining intraday-credit were fixed at the start of each day. They are revaluated daily with effect from the time when Kronos opens at 7:00 a.m. Changes in collateral value due to transfers to and from the accounts were very limited in January 2008 during Kronos' opening hours.

The timing of payment settlements in the system is fairly stable, as illustrated in figure B.1. With a few exceptions, more than 90 per cent of all payments were settled before noon in 2007. In the last hour of Kronos' opening hours, system activity was very limited.

[26] More information on the network topology of Kronos can be found in Rørdam and Bech (2008).

Figure B.1 **Time profile of interbank payments in Kronos, 2007**



Note: Figure shows accumulated share of daily settlement at specified times during Kronos' opening hours.

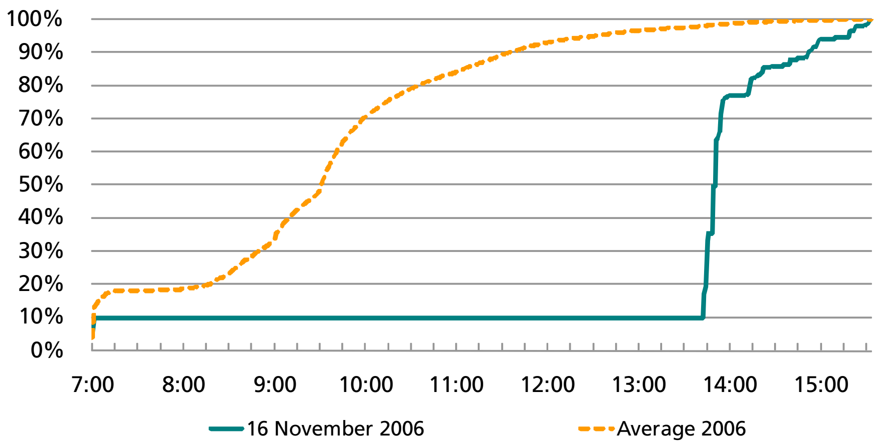
The data set satisfied the needs of our simulations. It shall be noted, however, that the exclusion of payments related to Danish ancillary systems prevented us from simulating incidents lasting more than one day (because net payments in the night settlement of the ancillary systems often cause a significant redistribution of liquidity between participants). This restriction should be removed in future research, as incidents lasting several days happen, albeit rarely.

Appendix C

System incident in Kronos 16 November 2006

The settlement of payments in Kronos was particularly disrupted by several system incidents in the second half of 2006 due to problems with the IT supplier's connection to SWIFT.[27] Most failures were of short duration and did not require special measures, but in order to prevent problems with CLS settlement, Danmarks Nationalbank in three cases had to initiate contingency procedures with respect to manual transmission of funds between participants. In connection with the most serious incident, on 16 November, which lasted 6½ hours, DKK 12 billion was transmitted between participants in order to ensure timely settlement of foreign-exchange contracts with a total principal of DKK 148 billion.

Figure C.1 **Time profile of interbank payments settled in Kronos on 16 November 2006**



Note: Figure shows accumulated share of daily settlement at specified times during Kronos' opening hours.

The disruption of payment settlement caused by the incident on 16 November 2006 was so severe that it had an impact on the Danish money market. Consequently, Danmarks Nationalbank extended its opening hours and gave access to extraordinary buy-back of

[27] See Danmarks Nationalbank (2006).

certificates of deposit. Soon thereafter, conditions in the money market normalised.

In the end, the impact of the incident was more or less limited to a delay in the settlement of payments, as shown in figure C.1. The figure also shows that when Kronos was operating again, from around 13:40 in the afternoon, the backlog of payments then released from the participants was quickly settled with almost 70 per cent of that day’s transactions settled within 20 minutes.

As a consequence of the incidents in the second half of 2006, a project was launched to improve the IT provider's SWIFT environment.

Incident at a large participant, March 2003

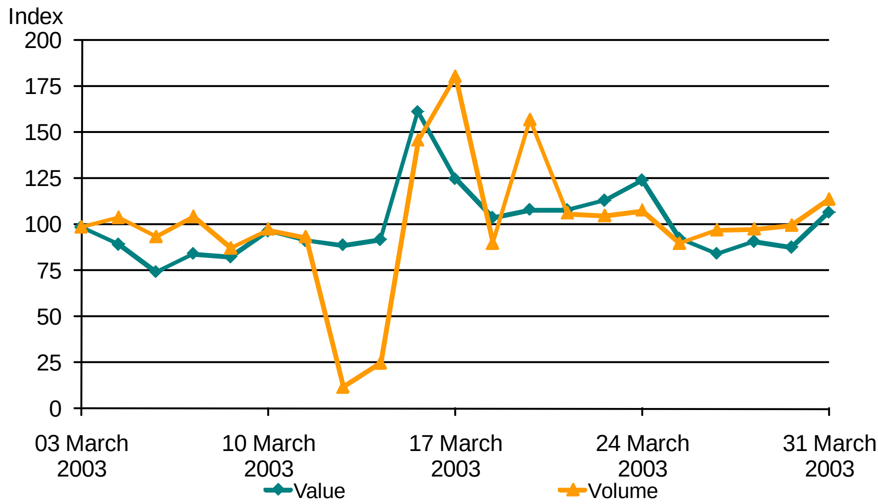
One of the largest participants in Kronos was hit by an incident in its IT systems in the afternoon on Monday 10 March 2003. The participant was subsequently unable to participate normally in payment systems in Denmark and abroad for several days.

In relation to Kronos, the incident severely impacted the participant’s capacity to send payments to other participants on Tuesday and Wednesday 12–13 March (see figure C.2). However, the contingency measures put in place ensured that the participant was able to send a number of large payments, so that the bank’s payments activity in terms of value was comparable with other banking days in March 2003.

On the days after normal operations were resumed, the participant hit by the incident had to send an extraordinarily large number of payments to other participants. Cleaning up after the incident entailed payments sent from the participant on Friday 14 March that, in terms of value, amounted to 60 per cent more than the participant’s average payments activity in Kronos in March 2003. Cleaning up was not completed before the following week and an extraordinary large number of payment orders were sent on 17 and 19 March 2003.

The incident had some impact on other participants, which experienced a slight decrease in their liquidity positions. This did not, however, significantly affect their behaviour in relation to payments in Kronos. Danmarks Nationalbank only had to provide a limited amount of extra liquidity to other participants because of the incident.

Figure C.2 **Settlement of payment orders from large participant affected by incident in second week of March 2003**



Chapter 3

Liquidity effects of a participant-level operational disruption in the Swiss Interbank Clearing System

Martina Glaser – Philipp Haene

3 Liquidity effects of a participant-level operational disruption in the Swiss Interbank Clearing System

Abstract

This paper shows that in an adverse scenario an operational disruption of a major participant in the Swiss large-value payment system, Swiss Interbank Clearing (SIC), can lead to significant systemic liquidity effects. Our analysis is based on simulation methods employing the BoF-PSS2 simulator developed by the Bank of Finland. We simulate an operational disturbance which prevents a major participant from inputting payments into the system. Due to a suspension of the participant's outgoing payments, liquidity can accumulate on its account and cause liquidity shortages for other participants and disrupt settlement of their payments. Our simulations show that under certain assumptions, the daily average systemic effect in SIC corresponds to 36 billion Swiss francs – or 22% of total payment value. However, the effect in terms of number of payments is much smaller. Subsequently, we examine the determining factors of the systemic liquidity effect such as participants' input behaviour and system liquidity levels. Our work highlights the importance of measures aimed at preventing operational disruptions or limiting their negative effects on other participants.

3.1 Introduction

Safe and efficient payment and securities settlement systems are a key prerequisite for the smooth functioning of an economy and an integral component of a stable financial system. One of the features of real-time gross settlement (RTGS) systems is that participants' liquidity management relies on the constant recycling of liquidity from incoming payments. This enables participants to effect payments with liquidity levels equivalent to only a small fraction of the total values settled. However, operational disruption of a participant during the day can interrupt this recycling mechanism as liquidity accumulates on the account of that participant and a 'liquidity sink' develops. This can lead to systemic effects,[1] as other participants lack the liquidity to settle their payments. Given the network of interdependencies, these systemic effects can spread to other connected systems.

Research on payment and securities settlement systems is typically hampered by the complexity of these systems, the dynamic behaviour of the participants and the vast amount of data, which make traditional econometric models difficult to apply. As a result, simulation methods have become a widely used tool. We use simulations – employing the BoF-PSS2 simulator developed by the Bank of Finland – to analyse the systemic impact of a participant-level operational disruption in the Swiss large-value payment system, Swiss Interbank Clearing (SIC). We further identify the main factors driving the size of the systemic effect and the measures taken to mitigate it.

SIC is one of the largest RTGS systems in terms of number of transactions, with 1.4 million transactions settled on an average day in 2008. Being linked to almost all the other payment and securities settlement systems in the country, it is the core of the Swiss financial market infrastructure. SIC also settles pay-ins and pay-outs in Swiss francs related to Continuous Linked Settlement (CLS), a multi-currency payment system for settling foreign exchange transactions. Therefore, systemic effects in SIC can lead to contagion effects in interdependent payment and securities settlement systems. Moreover, these other systems can be a source of liquidity shocks in SIC.

This paper is organised as follows: Section 3.2 reviews the results of similar studies in other countries. Section 3.3 highlights the key characteristics of the SIC system. Section 3.4 provides a conceptual overview of the liquidity effects of an operational disruption in payment systems. This is followed by a description of the simulation methodology in Section 3.5. In Section 3.6, the results of the simulations for the systemic effects of operational disruptions are presented and compared to other studies. Section 3.7 discusses the

---

[1] For the purpose of our study, we define systemic risk as the possibility that a single participant's inability to meet its SIC obligations due to operational problems will render other participants unable to meet their SIC obligations when due. Systemic effects are measured by the value and number of payments from a participant which are not settled in SIC due to liquidity shortages stemming from an operational problem at another participant. The daily average systemic effect is then calculated as the average of the systemic effects for every day over the 18 business days in May 2004 used in the simulation. Note that according to this definition, systemic effects would not necessarily lead to a systemic crisis or threaten financial stability. The decisive feature of systemic effects in our study is that due to liquidity shortages stemming from an operational problem at one participant, other participants cannot settle their own payments in SIC.

different factors that drive the systemic effect for SIC. Special attention is devoted to participants' input behaviour. Section 3.8 presents the measures that can be taken in case of operational disruptions. Section 3.9 concludes and lists areas of interest for future research.

3.2 Literature review

Bedford, Millard and Yang (2004) were among the first to apply simulation techniques to study systemic effects of participant-level operational disruptions. They concluded that for the CHAPS Sterling payment system in the UK systemic effects are unlikely to occur given the very high liquidity levels in the system (150% of upper bound liquidity)[2] and the other banks' fast reaction time (10 minutes) on which their simulation is based. A significant systemic effect was only identified in a theoretical scenario where three CHAPS Sterling settlement banks were hit by an operational disruption when effective liquidity was below the upper bound.

Subsequently, various studies have investigated systemic effects of an operational disruption in interbank payment systems. Some of these found only very minor systemic effects, for example Bech and Soramäki (2005) for the US Fedwire, and McVanel (2005) and Ball and Engert (2007) for Canada's large-value transfer system. Others found moderate systemic effects. For example, Ledrut (2007), Mazars and Woelfl (2005) and Hellqvist/Snellman (2007), making certain assumptions, found moderate indirect liquidity effects in the Dutch and French payment system and in the Finnish equities settlement system.

3.3 Swiss Interbank Clearing (SIC)

SIC plays a pivotal role in the Swiss financial market, as it settles all large-value payments and a large number of retail payments in Swiss francs. Linked to almost all other payment and securities settlement systems in the country, it is the core of the Swiss financial market infrastructure. SIC settles money market transactions, the cash-leg of

[2] For a description of the concept of upper bound liquidity see for example Koponen and Soramäki (1998).

securities transactions, and CLS related pay-ins and pay-outs in Swiss francs. Its safe and efficient functioning is critical for implementation of the Swiss National Bank's monetary policy.

The system is operated by SIX Interbank Clearing AG on behalf of the Swiss National Bank (SNB), and transactions are settled on accounts at the SNB. It is an RTGS system with central queues, and processes payments according to priorities and the first-in first-out rule. Through a link with the SECOM securities settlement system, SIC guarantees settlement of securities transactions according to the principle of delivery-versus-payment (DvP). SIC operates around the clock on bank working days. Payments can be entered at any time, and up to five days in advance. Settlement takes place for 23 hours of every day, starting at around 5 pm on the calendar day before the value date and continuing until around 4.15 pm on the value date. To ensure smooth functioning of SIC, the SNB provides intraday liquidity through its repo facility. Of the approximately 330 participants, the two major participants account for roughly a half of the transaction values settled in SIC.[3]

3.4 Overview of liquidity effects of a participant-level operational disruption in interbank payment systems

The risks arising in interbank payment systems can be broadly grouped into credit risk and liquidity risk. Credit risk is the risk that a party will be unable to fully meet its financial obligations within the system on the due date or at any time in the future. Liquidity risk is the risk that a party will have insufficient funds to meet its financial obligations when due, although it may be able to do so at a later date. Operational problems and legal uncertainties can cause credit and liquidity risk. Central banks are especially concerned with systemic

[3] Figures for 2004. For a more detailed description of SIC, see Heller, Nellen and Sturm (2000).

risks,[4] as these can endanger financial stability, which, usually, is an explicit or implicit objective of the central bank. Credit and liquidity risks, operational disruptions, and legal uncertainties can be sources of systemic risk, if the failure or delay of a participant to meet its obligations, or a disruption in the system itself, will cause other system participants or financial institutions to be unable to meet their obligations as they fall due. Such a failure may cause widespread liquidity or credit problems and thus threaten the stability of the financial system or even the economy as a whole. SIC – being an RTGS system – eliminates credit risk in the settlement process, as settlement is final within the day. Liquidity risk is mitigated by providing easy access to intraday liquidity from the central bank and through a gridlock-resolution mechanism.[5]

In recent years, due to events such as the 'Year 2000' date change, the terrorist attacks on financial centres like New York and London, and the threat of global pandemics like SARS or avian flu, increasing attention has been paid to potential systemic effects of operational disruptions. As a result, operators of payment systems have updated their business continuity plans and tightened operational requirements for critical system participants. For example, in Switzerland, an industry group has published recommendations for improving business continuity planning which were subsequently adopted in the self-regulatory best practices published by the Swiss Bankers Association.[6]

In the event that a participant of an interbank payment system suffers an operational disruption, two effects can be distinguished. First, other participants will not receive payments from the disrupted participant and may cancel payments to the participant in question after a certain period of time (direct effect); and second, other participants may not be able to settle their own payments due to liquidity shortages caused by the liquidity sink stemming from the failing participant. The latter effect is typically referred to as the indirect or systemic effect of an operational disruption. Systemic

---

[4] The ECB and Eurosystem (2008) defines systemic risk as 'the risk that the inability of one participant to meet its obligations in a system will cause other participants to be unable to meet their obligations when due, with possible spillover effects such as significant liquidity or credit problems that may threaten the stability of or confidence in the financial system.' Note that our definition in the context of this study (see footnote 1) focuses on the inability of participants to meet their own obligations due to an operational problem of another participant. It ignores whether this has the potential to threaten the stability of the financial system.

[5] For a detailed discussion of risks in large-value payment systems see CPSS (2005).

[6] See Industry group (2006) and Swiss Bankers Association (2007).

effect refers here to the second-round effects within the payment system and need not imply that financial stability would be compromised by this indirect effect. While the direct effect can have a significant impact on other participants, it can be calculated in a straightforward manner. In contrast, the systemic effect of a participant's disruption is less obvious, as it depends on the dynamics of the settlement process and participants' behaviour. We therefore estimate the systemic effect by simulating 18 business days in May 2004. We show that in an adverse scenario the daily average systemic effect in SIC corresponds to 36 billion Swiss francs – or 22% of total payment values. However, in terms of the number of payments, the effect is much smaller.

3.5 Simulation methodology

Data sample

Our simulations are based on SIC transaction data for May 2004. The data sample covers 18 business days and 12,950,000 payments totalling 2,983 billion Swiss francs. The very large transaction volumes can be explained by the fact that SIC settles both large-value and small-value payments. In May 2004, transaction volumes and values were equal to the averages for the year 2004, and they exhibit a typical monthly pattern with a peak towards the end of the month. This provides an indication – though not certainty – that May 2004 was a representative month for SIC in that year.

Differences in the settlement algorithm

Using the simulator version BoF-PSS2 1.2.0, it is possible to mimic the functionality of an RTGS with central queues. However, there are some specific settlement characteristics of SIC which were not replicated in our simulations. These are outlined in Table 3.1.

Table 3.1 **Differences between SIC settlement algorithm and its simulation**

	Actual SIC	SIC Simulation
Queue release algorithm and packet building	Payments are queued and processed according to payment priority (first criterion) and input date and time of payment (second criterion). Due to huge numbers of SIC payments and processing capacity limits, one or several accounts can include queued payments even with sufficient available funds. In such case, SIC selects the queue to be settled first according to input date and time of first payment, regardless of priorities. Once a queue is selected for settlement, the algorithm continues to settle payments from that queue as long as funds are available, the priority is the same, the maximum packet size of 100 payments is not reached, and the next payment is not more than 50 seconds younger than the first payment (based on input time).	Payments are queued and processed according to payment priority (first criterion) and input date and time of payment (second criterion). All queued payments are settled as long as sufficient funds are available. Since capacity limits are not simulated, there will never be accounts with queued payments when sufficient funds are available.
Bilateral off-setting	If no payment is settled within 15 seconds, the system initiates a bilateral off-setting mechanism.[7]	No bilateral off-setting mechanism.
CLS accounts	A CLS settlement member for Swiss francs can have a special CLS account in SIC to initiate such payments, which prevents queues in its regular account from blocking time-critical pay-ins to CLS.	Regular accounts and special CLS accounts are combined in a single account for simulation purposes.[8]
Opening hours	For value dates after weekends and bank holidays, a SIC settlement day lasts for more than 24 hours.	Since the simulator version used could only handle settlement days of up to 23:59 hours, the SIC opening was delayed for value dates after weekends and bank holidays. Thus payments were queued in our simulation, although they would have been settled immediately in practice.

[7] The bilateral off-setting mechanism for solving gridlocks is seldom used (in July 2007 – July 2008, it was used 142 times, with a total value of 8,727 million Swiss Francs). For a description of the bilateral off-setting mechanism, see Sturm (2000).

[8] This explains why in Figures 3.1 and 3.2 (which show an aggregated view of the CLS account and the regular account) the participants may exhibit queued payments although sufficient funds are available for settlement. In these cases, the funds are on one account (eg the CLS account) while the queued payments are on the other account (eg the regular account).

We cannot determine conclusively the extent to which these differences distort our findings on the SIC liquidity and the systemic effects of operational problems. However, test simulations indicate that our simulations should mimic liquidity effects in SIC without major distortion.[9]

Simulation assumption

For our simulations we broadly follow the methodology developed by Bedford, Millard and Yang (2004) in their assessment of systemic risk in CHAPS Sterling. We simulate operational disruptions that prevent one of the two major participants from inputting payments into SIC from a certain time onward until the end of the settlement day. Queued payments that are inputted by the affected participant prior to the disruption are settled. We assume that the other participants will cancel payments to the disrupted participant two hours after the disruption has occurred. During these two hours, the onset of a liquidity sink is possible, as the disrupted participant is receiving payments but is unable to initiate new payments. We also assume that the disruption occurs at the moment when the theoretical liquidity sink is largest, given the scenario described above. Algebraically, we find the largest theoretical liquidity sink $LS_{i,t}$ (and hence the moment when the participant-level disruption occurs) by maximising the expression

$$LS_{i,t} = B_{i,t} + \sum_{t}^{t+120} IP_{i,t} - Q_{i,t}$$

for participant i, where $B_{i,t}$ denotes his account balance at time t, $\sum_{t}^{t+120} IP_{i,t}$ is the value of incoming payments of participant i over the next 120 minutes, and $Q_{i,t}$ is the value of payments in the queue of i at time t.

The timing of the largest theoretical liquidity sink depends on participants' input and settlement behaviour. We refer to theoretical liquidity sink because it does not account for liquidity restrictions on the other participants. According to our calculation and ignoring liquidity restrictions on the other participants, the potential liquidity trapped on the account of the failing participant lies between 7 and 25 billion Swiss francs.

[9] However, the delay indicators could be significantly impacted by these differences. We did therefore not rely in our analysis on the delay indicators generated in the simulations.

While our methodology is similar to that of Bedford, Millard and Yang (2004), a few important assumptions are different. In contrast to CHAPS Sterling, SIC uses centralized queues. This means that payment orders in the queue of the disrupted participant will still be settled after the disruption occurred. Furthermore, we also differ in our assumptions on reaction time for other system participants to cancel payments to the disrupted participant. While the UK simulation assumes that the other participants stop their payments to the disrupted participant 10 minutes after the incident, we assume that participants stop sending and executing payments to the disrupted participant only two hours after the failure occurs. Our assumption regarding the non-disrupted participants' behaviour is motivated by their input behaviour in past incidents in SIC. Based on these incidents, we assumed for our simulations that in the event of an operational disruption, the disrupted participant continues to receive payments until the operational failure is communicated and the extent of the operational problem becomes evident.

This might however not be the case for other payment systems. Amanuel and Conover (2005) identify historical disruptions by looking at unusual delays between payments. They then evaluate the number and value of payments received by these potentially disrupted participants and find some evidence that temporarily disrupted banks receive less payment values. This suggests that non-disrupted participants react to temporary operational disruptions of a participant by quickly and significantly altering their payment behaviour. This finding is in line with McAndrews and Potter (2002), who estimate a reaction function where participants' payment sending behaviour is related to payment receipts. Inspired by these studies, Ledrut (2007) simulates systemic effects of participant-level disruptions considering different scenarios for behavioural changes.

We consider it important to distinguish two different types of behaviour changes for non-disrupted participants. First, behaviour changes of non-disrupted participants can reflect a deliberate decision to delay payments because of uncertainty about a disrupted participant and its ability to pay its obligations. Second, delays in payments might not be based on an explicit decision but may instead reflect liquidity effects because of a lack of incoming funds from other participants. While the first type of behaviour change would probably be reflected in input behaviour, the second type would be reflected in available liquidity and thus in the number and value of payments actually settled. We reviewed the SIC data from historical incidents to evaluate these behaviour changes. This was facilitated by the fact that the SIC data includes both input and settlement times of payments. Looking at past incidents, we could not identify behaviour changes for both effects. However, it must be stressed that we had to rely on a very small sample of historical participant-level disruptions. Also, these historical incidents tended to be of a much smaller scale than the disruptions assumed in our simulations.

3.6 Simulation results

Our simulations show significant direct effects from a disruption of a major participant in SIC: on average, 24% of payment values and 4% of the number of payments (see Table 3.2). This comes to a daily average of approximately 32,000 payments totalling 40 billion Swiss francs. These payments are either not inputted into or are deleted from the payment system queue following a participant's operational disruption. As noted in Section 3.5, we assume that the disrupted participant immediately stops sending payment orders, whereas the other participants take two hours to react and stop sending payment orders to the disrupted participant or deleting such orders from the queue.

Table 3.2 **Direct and systemic effects of an operational disruption of a major SIC participant (daily average for 18 business days in May 2004)**

	Number of transactions		Value of transactions in million CHF	
May 2004	Daily average	% (minimum to maximum)	Daily averages	% (minimum to maximum)
Settled transactions	661 000	92 (85 to 98)	89 100	54 (33 to 93)
Unsettled transactions – direct effect	32 000	4 (1 to 7)	40 400	24 (4 to 44)
Unsettled transactions – systemic effect	26 000	4 (0 to 9)	36 200	22 (1 to 37)
Total	719 000	100	165 700	100

Also the systemic (indirect) effect of such a disruption can be large in an adverse scenario. In terms of payment values, an average of 22% – or 36 billion Swiss francs – would not be settled due to systemic effects.[10] However, in terms of the number of payments, the systemic effect is much smaller, affecting only 4% of transactions. This substantial difference is due to the fact that many retail payments of small value are settled before the disruption is assumed to occur. On average, around one third of all participants would be affected.

It should however be stressed that these results for the systemic effect are based on rather extreme assumptions and ignore the possibility of participants' access to additional liquidity from the central bank or other sources in the market. Our assumptions therefore most likely resemble a crisis scenario. Depending on the actual operational disruption and taking into account the measures taken to mitigate liquidity and operational risks (see Section 3.8), actual systemic effects would typically be much smaller.

Irrespective of the underlying assumptions, significant systemic effects are limited to disruptions of the two major SIC participants; simulating the disruption of other, smaller participants results in minor systemic effects only. The results are therefore only valid for the two major participants in SIC. Also, the size of the systemic effect varies considerably among major participants and even among different days for the same participant. The systemic effect, in terms of payment values, ranges from 1% to 37%, depending on the day and disrupted participant.

Comparing our results with other simulations, we find larger systemic effects from a participant-level operational disruption in SIC compared to similar studies conducted so far. For example, Bedford, Millard and Yang (2004) find only minor systemic effects from operational disruption of a participant in CHAPS Sterling. However, their simulations are based on different assumptions (see Section 3.5). In another study, Ledrut (2007) concludes that systemic effects of an operational disruption of a major participant in the Dutch large-value payment system TOP are limited, due to relatively high liquidity levels in the system. While Mazars and Woelfl (2005) and Hellqvist and Snelmann (2007) find some systemic effects, they are clearly smaller than in SIC. The same is true for the studies conducted for the Canadian large-value transfer system by Ball and Engert (2007) and McVanel (2005), and for the US Fedwire system by Bech and Soramäki (2005). While the differences in systemic effects can to an extent be attributed to simulation assumptions, there is still a residual of difference that must be explained by system-inherent factors.

[10] For the calculation method for the systemic effect, see footnote 1.

Therefore, in the next Section, we will evaluate the main drivers of the large systemic effects in SIC.

3.7 Factors driving systemic effects

There are various factors that influence the systemic effects of a participant-level operational disruption in a payment system. In the following paragraphs, we discuss these factors in light of our simulation results for SIC.

Liquidity turnover

One of the main characteristics of SIC is its relatively high liquidity turnover. On an average day in May 2004, 165.7 billion Swiss francs were settled with only 11.6 billion Swiss francs of liquidity, which equals 7% of settled values. This means that liquidity in SIC was turned over more than 14 times in a single business day.[11] While this means that participants can settle their payments with low liquidity levels and can therefore invest more funds in assets that may provide higher returns, it also has drawbacks. In particular, with a low level of liquidity in the system, it is likely that the emergence of a liquidity sink will drain other participants' accounts faster and lead to larger systemic effects.

Participant structure

Another factor influencing the potential size of the systemic effect is the payment system's participant structure, which is highly concentrated in the case of SIC. As mentioned above, the two largest participants account for roughly 50% of all payments, in terms of value. If one of these major participants faces an operational problem,

[11] Liquidity levels in the large-value payment systems in Norway and Sweden amount to around 70% and 20% respectively of total settled values (see Enge and Overli, 2006 and Sveriges Riksbank, 2003). However, one should not draw conclusions from these turnover figures on the liquidity efficiency of the systems. Differences between systems may reflect differing reserve requirements or liquidity supply costs for participants. For example, February 2009 figures for liquidity turnover in SIC show that it has dropped to only 4 times, as the Swiss National Bank has cut interest rates and substantially increased liquidity available to SIC participants.

the consequences are more severe than in a less concentrated system. Indeed, our simulations indicate that only the two largest participants have the potential to cause significant systemic effects. The disruption of any other participant will have only minor systemic effects. This finding is supported by Nier, Yang, Yorulmazer and Alentorn (2008), who conduct simulation experiments to examine contagion effects in a theoretical banking system with interbank exposures. They conclude that, ceteris paribus, the higher the concentration in the banking system, the more vulnerable it is to systemic risk. This result holds true irrespective of the size of a given shock.

Settlement algorithm

Furthermore, the size of the potential systemic effect is influenced by the system design. As described above, SIC is an RTGS system with only limited liquidity saving mechanisms. Systems with continuous offsetting mechanisms may be able to settle more payments in circumstances when liquidity is scarce than those without such features. However, as these mechanisms are also used in normal situations, participants might anticipate their use and lower their precautionary liquidity levels, so that the systemic effect is not necessarily reduced.

Participants' input behaviour

Our results also suggest that the input behaviour of a participant has a major impact on the size of the systemic effect. Looking at our May 2004 data sample, we find evidence of different input behaviour for the two largest participants, but also within the same participant on different weekdays.

For example, figures 3.1 and 3.2 show the stylised input behaviour of the two major participants, labelled A and B. A enters most of its payment orders (in term of values) early in the morning (at around 7.30am). This leads to a large queue in the morning, which is reduced over time because SIC settles the queued payment orders automatically as liquidity becomes available on A's account (see figure 3.1). B, however, seems to manage its payment input to avoid the build-up of queues in SIC. It only enters payment orders if settlement liquidity is available. Therefore, its queued payment values are typically low (see figure 3.2). This difference in input behaviour has implications if a participant faces an operational disruption and is unable to send new payment orders to SIC. In the case of A, the queued payments act as a buffer, as liquidity is automatically recycled. The potential liquidity sink on the disrupted participant A's account will be smaller and the systemic effect will be reduced. By contrast, if participant B is unable to send payment orders to SIC, this will be immediately felt from a liquidity perspective as a liquidity sink accumulates and drains other participants' liquidity levels.

Figure 3.1 **Participant A – Early morning input of payment orders (stylised)**

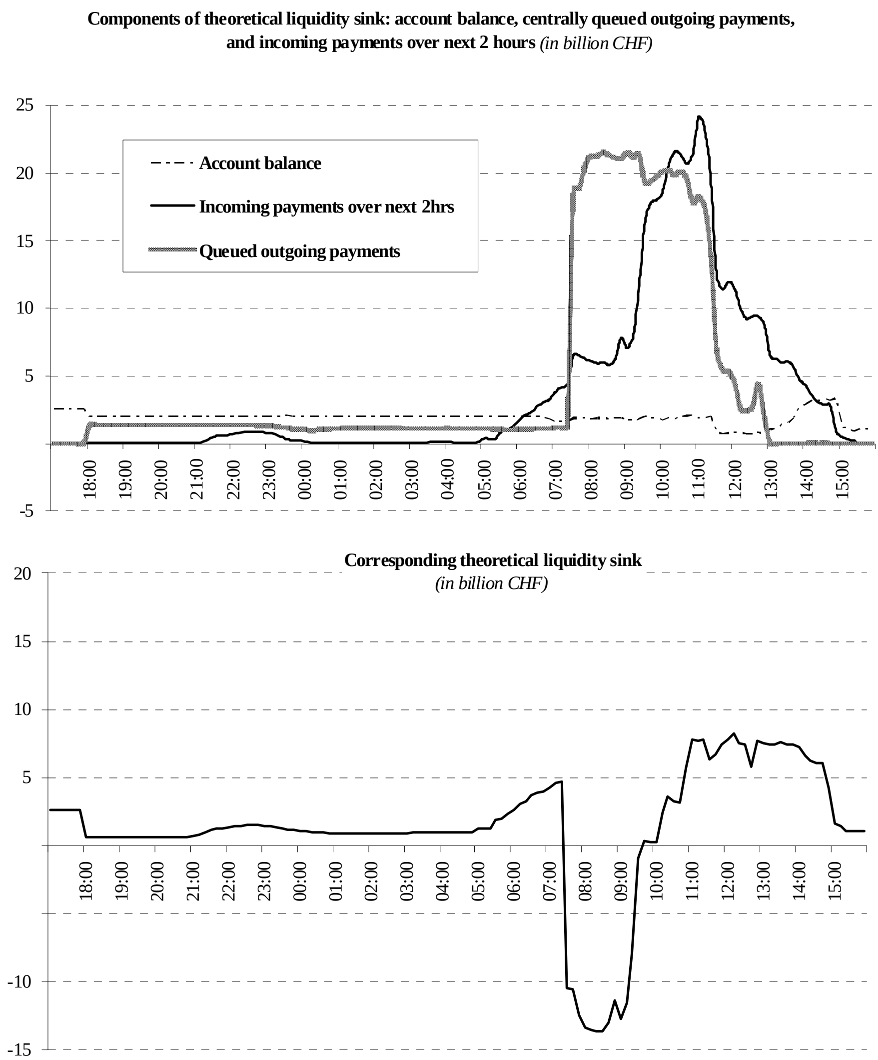
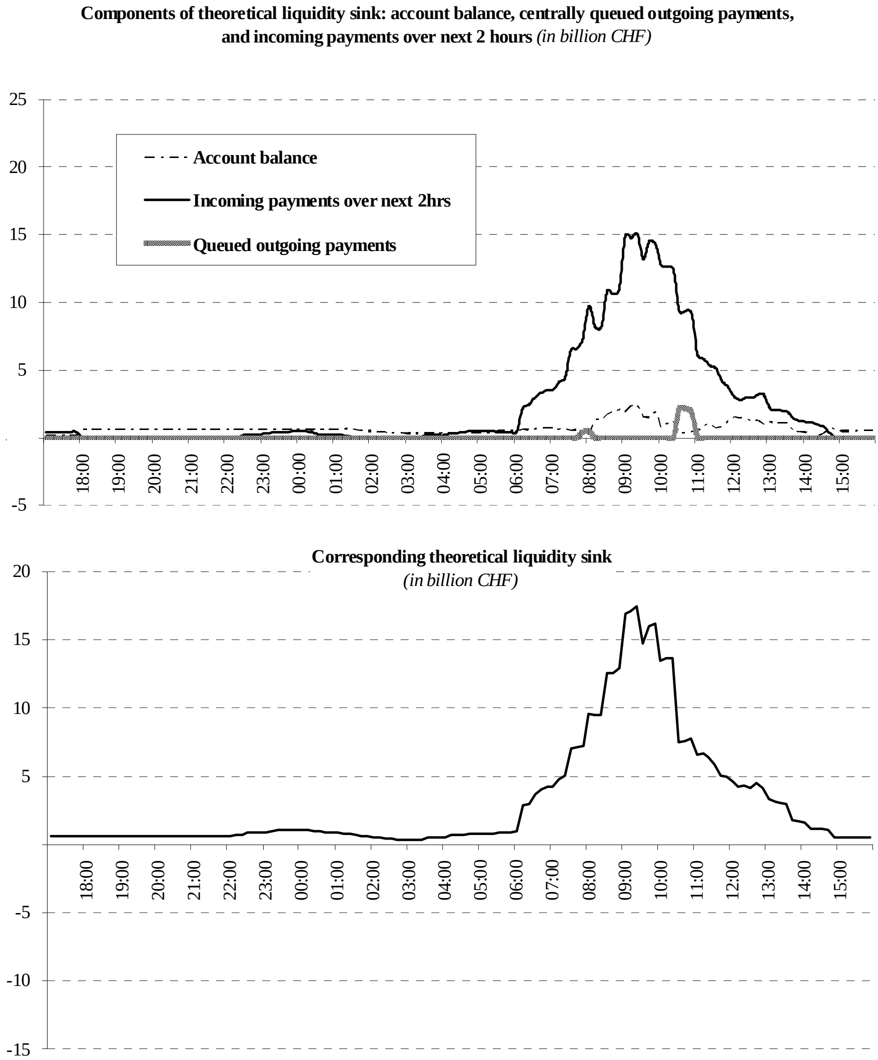


Figure 3.2 **Participant B – Continuous input of payment orders in the course of the day (stylised)**



Also, we find that the input behaviour of the major participant A varies depending on the weekday. Participant A receives high-value payments in the morning, typically between 8am and 11am. Hence, this is a time when a large liquidity sink could arise. For this participant, a small difference in input behaviour has a significant impact on the size of the potential systemic effect arising from an operational disruption. This is because we assume the disruption occurs when the potential liquidity sink is largest. On Mondays and on days after bank holidays, this participant inputs the majority of its payments (in terms of values) before 7.30am (see Figure 3.1). On these days, the systemic effect of an operational failure of this participant is relatively small. This can be explained by the fact that the value of outgoing payments already stored in the central queue is larger than the value of incoming payments. Therefore, on Mondays and after bank holidays, the biggest theoretical liquidity sink for participant A arises rather late during the settlement day, when a large share of payments has already been settled. On the contrary, the effect of an operational problem of participant A on other weekdays is very large in terms of value. On these days, it inputs its payment orders only after 8am and consequently, very large liquidity sinks can arise if a disruption occurs just before 8am. This small difference in input behaviour leads to large variations in the size of the systemic effect, given our assumptions.

These examples also suggest that under certain circumstances queues can be beneficial for a payment system. Queues in payment systems might be seen as a consequence of a shortage of liquidity which then leads to settlement delays. The conclusion would then be that queues should be avoided. Our results show, however, that central queues which reflect the early input of payment orders act as a potential shock-absorber for the liquidity effects of a participant-level disruption. For this reason, central queues arising early in the settlement day can have positive effects. Of course, central queues arising late in the day can be a source of liquidity risk, as settlement failures or even minor operational disruptions can have significant consequences on the liquidity planning of other participants and can lead to short-term liquidity needs.

In general, we conclude that early input of payment orders into the system (even if queued centrally and not immediately settled) reduces the theoretical liquidity sink, compared to late input of payments.[12] In the event that a participant with early input experiences an operational disruption and is unable to send new payment orders to the system, the queued payment orders would help recycle liquidity. This would prevent the emergence of a systemic effect or at least limit its size. If, however, a participant inputs payments only late in the day, an operational disruption occurring before that will be felt immediately in the settlement process, as no payments are queued. In such cases, the

[12] Of course, it must be noted that participants' leeway in scheduling payment orders is limited. Payment orders might be arriving late from customers or internal sources, preventing early input of such payments.

time span before a liquidity sink builds up will be shorter and the liquidity trapped in the liquidity sink will be larger. Variation in the daily input behaviour of major participants largely explains the wide range of potential systemic effects experienced in our simulations (between 1% and 37%, depending on the day and disrupted participant).

Simulation assumption

Several assumptions were made for our simulations which significantly affect the results. For example, we assumed that the disruption of a major participant takes place at the time when the theoretical liquidity sink is highest. Further, we supposed that non-disrupted participants only cancel payments to the affected participant two hours after the disruption has occurred. This last assumption is based on anecdotal evidence from real temporary disruptions of participants in SIC and the subsequent behaviour of other participants. If the reaction time of the non-disrupted participants were shortened, this would reduce the potential systemic effect.

Our assumption for the timing of the disruption is the moment when the theoretical liquidity sink is highest. However, this need not correspond to the largest potential systemic effect. To identify the moment when a disruption triggers the largest systemic effect, one needs to take into account not only the liquidity sink, but also the payment distribution over the day and between participants. Therefore, the largest systemic effect cannot be derived directly but would require a multitude of simulations for every participant for every settlement day on a trial-and-error basis.

3.8 Measures to mitigate systemic effects

As shown, assuming an adverse scenario the operational disruption of a major participant can cause significant systemic effects in SIC. This highlights the pivotal role of sound business continuity measures – not only at system-level but also at participant-level – and the importance of adequate incentives and instruments for all involved parties to mitigate systemic effects.

First, preventive measures have been taken to minimise the likelihood of a prolonged operational disruption of a participant in SIC. An industry group has established recommendations for the maximum down-time of critical participants in SIC.[13] In its report, this industry group suggested that critical participants should be able to resume operations within four hours after the loss of a key building, including the staff working in the building.[14] These recommendations have been integrated in self-regulatory best practices published by the Swiss Bankers Association.[15] Specific measures taken to adhere to these best practices include the establishment of redundant data centres and backup communication networks to access the payment system.

Second, there is a variety of incentives and instruments to reduce systemic effects should a prolonged disruption occur. As illustrated above, early input of payments can reduce the size of the potential systemic effect. In SIC, early input is encouraged by a progressive fee structure. Furthermore, the end-of-day cut-off time can be postponed, should a temporary operational disruption occur. Participants also have access to intraday liquidity from SNB, based on a wide range of collateral. Also, payments on behalf of the affected participant can be entered by SNB, which helps to redistribute the liquidity among participants. Further, SNB has the technical possibility to initiate an individual clearing stop for a disrupted participant, which immediately stops all outgoing and incoming payments to and from this participant. In addition, there are facilities for physical input of data via tapes should the telecommunication infrastructure be unavailable. Finally, an interbank alarm and crisis organisation exists to coordinate industry reactions.

Our analysis of the driving factors of systemic effects also highlights additional measures which might prove to be even more effective in dealing with disruptions. These include:

- establishing bilateral or multilateral sender limits to reduce the size of a liquidity sink arising from an operational disruption;
- influencing the input behaviour of critical participants, eg by establishing through-put requirements. These would require major participants to settle a certain percentage of payment obligations before a given time on the settlement day;
- adapting the settlement mechanism in SIC to include additional liquidity optimisation algorithms; and

[13] For defining critical participants, a threshold of a 5% (or slightly below) value share in SIC was used.
[14] See Industry Group (2006).
[15] See Swiss Bankers Association (2007).

- enhancing the existing interbank alarm and crisis organisation to reduce reaction times of other participants should a disruption occur. One concrete option would be to include communication network operators in the alarm and crisis organisation.

3.9 Conclusions and outlook

Our simulations suggest that in an adverse scenario the systemic liquidity effect of an operational disruption of a major participant in SIC can be very large. On an average day in May 2004, 22% of payment values – ie a total of 36 billion Swiss francs – would not be settled and about one third of the other participants could be affected. As SIC is also closely connected to other Swiss payment and securities settlement systems as well as to international systems, these systemic effects could lead to further contagion effects if liquidity shocks are transferred to other systems.

We find that some of the factors accounting for the large systemic effect are participants' input behaviour, the relatively low liquidity levels in SIC and the high concentration ratio of its participants. Since pending payment orders can act as a shock absorber for liquidity effects, our results suggest that under certain circumstances queues can be beneficial for a payment system. To the extent that queues are managed centrally in the payment system and reflect participants' early inputs of payment orders, the build-up of such queues – at least early in the day – can act as a liquidity shock absorber in case of operational disruptions. In SIC, the fee structure encourages early input of payment orders into the payment system, even if they are not immediately settled but initially queued in SIC's central queue.

There are two important caveats to our analysis. First, the potentially large systemic effects in SIC underline the importance of taking measures to mitigate liquidity and operational risk in SIC. As these measures were not incorporated in our simulations, our analysis identifies the potential size of systemic effects in SIC in a rather extreme scenario. For example, we ignore the possibility of participants having access to additional liquidity from the central bank and assume that compensating liquidity cannot be found from other sources in the market. Our assumptions therefore resemble a crisis scenario. Depending on the actual operational disruption and taking into account the measures taken to mitigate liquidity and operational risks, actual systemic effects would typically be much smaller.

Second, our simulations only allow for one behaviour change, which is that participants cancel payments to the affected participant two hours after a disruption has occurred. Other behaviour changes – for example faster reaction by other participants to the disruption or re-prioritisation of certain payments – are not considered.

Looking forward, an intriguing extension of our analysis would be to vary the assumed reaction behaviour of non-disrupted participants. For example, the measures already taken to mitigate systemic effects in SIC (which are highlighted in Section 3.8) could be incorporated in our simulations. Also, similarly to Ledrut (2007), the assumptions on participants' behaviour changes in response to a participant-level disruption could be varied to analyse the systemic effects. Finally, the usefulness of potential new measures could be evaluated. Specifically, the work of Mazars and Woelfl (2005) suggests that bilateral limits can be a powerful instrument for containing systemic effects of a participant-level disruption in interbank payment systems.

References

Amanuel, D – Conover, D (2005) **Operational Disruptions in Fedwire: Simulating Liquidity Needs and Understanding Counterparty response.** Presentation at the BoF-PSS Seminar, August 25.

Ball, D – Engert, W (2007) **Unanticipated Defaults and Losses in Canada's Large-Value Payments System, Revisited.** Bank of Canada Discussion Paper 2007-5.

Bech, M – Soramäki, K (2005) **Systemic risk in a netting system revisited.** In: Leinonen, H (Ed.) 2005.

Bedford, P – Millard, S – Yang, J (2004) **Assessing operational risk in CHAPS Sterling: a simulation approach.** Bank of England, Financial Stability Review, June 2004, Issue 16, 135–143.

Committee on Payment and Settlement Systems (CPSS) (2005) **New developments in large-value payment systems.** BIS/CPSS, May 2005.

ECB/Eurosystem (2008) **Glossary of terms related to payment, clearing and settlement systems.** European Central Bank, September 2008, available on www.ecb.int.

Enge, A – Overli, F (2006) **Intraday liquidity and the settlement of large-value payments: a simulation-based analysis.** Economic Bulletin, Norges Bank, 01/2006, 41–47.

Heller, D – Nellen, T – Sturm, A (2000) **The Swiss Interbank Clearing System.** Finanzmarkt und Portfolio Management, Volume 14(4), 413–431.

Hellqvist, M – Snellman, H (2007) **Simulation of operational failures in equities settlement.** In Leinonen, H (Ed.) (2007).

Industry Group (2006) **Business Continuity Planning in the Swiss Financial Centre.** January 2006, available on www.snb.ch.

Koponen, R – Soramäki, K (1998) **Intraday Liquidity Needs in a Modern Interbank Payment System – A Simulation Approach.** Bank of Finland Studies, E:14, 1998.

Ledrut, E (2007) **How can banks control their exposure to a failing participant?** In Leinonen, H (Ed.) (2007).

Leinonen, H (Ed.) (2005) **Liquidity, risks and speeding payment and settlement systems – a simulation approach.** Bank of Finland Studies, E:31, 2005.

Leinonen, H (Ed.) (2007) **Simulation studies of liquidity needs, risks and efficiency in payment networks.** Bank of Finland Studies, E:39, 2007.

Mazars, E – Woelfl, G (2005) **Analysis, by simulation, of the impact of a technical default of a payment system participant.** In Leinonen, H (Ed.) (2005).

McAndrew, J – Potter, S (2002) **Liquidity effects of the events of September 11, 2001.** Federal Reserve Bank of New York, Economic Policy Review, November 2002.

Nier, E – Yang, J – Yorulmazer, T – Alentorn, A (2008) **Network models and financial stability.** Bank of England, Working Paper No. 346, April 2008.

Sturm, A (2000) **Bilaterales Offsetting.** ClearIT, 2000 (13), 10–11.

Swiss Bankers Association (2007) **Recommendations for Business Continuity Management (BCM).** November 2007, available on www.swissbanking.org.

Sveriges Riksbank (2003) **Financial Stability Report.** 2003(2), 43–54, Stockholm.

McVanel, D (2005) **The Impact of Unanticipated Defaults in Canada's Large Value Transfer System.** In Leinonen, H (Ed.) (2007).

Chapter 4

Operational disruption and the Hungarian real time gross settlement system (VIBER)

Ágnes Lublóy – Eszter Tanai

4 Operational disruption and the Hungarian real time gross settlement system (VIBER)

Abstract

Central bankers wish to ensure worldwide that large-value transfer systems exhibit sufficiently robust levels of operational resilience. We focus on the operational resilience of the Hungarian real time gross settlement system, known as VIBER. The goal of the research is the quantitative assessment of the ability of the system to withstand the technical default of one or two systemically important participants. Altogether six plausible scenarios were formed, three entire-day incidents and three incidents involving less time (part-time incidents). When entire-day incidents, assuming no back-up options, and behavioural reactions were simulated, the disturbance of the payment system was severe. By international standards, the proportion of unsettled transactions was very high, which can be explained by the high concentration of debit turnover, the liquidity scarcity of the most active VIBER participants, and the structure and size of the money markets. We also shed light on the importance and potential efficacy of back-up options: the shock-absorbing capacity of the system improved significantly when back-up procedures were included. The impact of behavioural reactions of technically non-defaulted participants was also considered. Blocking payments to the stricken bank raised the value of unsettled payments, which is the price of isolating the shock and privileging payments sent to other participants.

4.1 Introduction

4.1.1 Motivation

The Hungarian real time gross settlement system (Valós Idejű Bruttó Elszámolási Rendszer or VIBER) is used mainly for settling large-value payments. The system can be considered critical infrastructure for those transactions that are to be settled in Hungarian forint. These transactions include mainly large-value financial market deals and other time-critical payments. The average daily turnover of VIBER amounts to approximately 10% of the annual GDP of Hungary. Given this scale of activity, if the system is inappropriately designed or poorly operated, it could expose participants to risk potentially large enough to threaten their day-to-day business activity. In extreme cases, the financial soundness of the participants and the stability of the system as a whole might also be threatened. A smoothly functioning large-value payment system is crucial to the efficiency of the financial markets. Besides the assurance of financial stability, central bankers have an additional interest in a resilient payment system, as it plays a pivotal role in the implementation of monetary policy.

A disruption to normal payment processing activity could result in the realisation of liquidity risk. If an operational problem involving a settlement bank prevents the bank from submitting payments to the system, then liquidity can accumulate on the defaulter's account (liquidity sink effect). As the bank is unable to redistribute liquidity in the system by submitting payments, the liquidity positions of the counterparties are also affected or even threatened. The counterparties may delay their payments or even worse: a lack of funds may render them unable to settle their payments.

Analysing the resilience of VIBER is part of the MNB's[1] payment systems oversight duties. Act LVIII of 2001 on the MNB defines as one of the basic tasks of the MNB the development of payment and settlement systems and monitoring of their activities in order to achieve sound and efficient operations and smooth money circulation (MNB, 2001). In line with this, the MNB has to assess all risks that might have an impact on the system overseen. If the central bank finds that there is high risk, it should take steps to eliminate or lessen the risk via proper risk management.

Accordingly, the goal of the current research is the quantitative assessment of the ability of the system to withstand certain types of operational shocks. We wish to shed light on the capacity of the system to function smoothly in the event of operational problems and highlight the mechanisms for mitigating the impact of such problems.

The paper is organised as follows. In section 4.1 we review the operational incidents in VIBER and outline the scope of the research. Section 4.2 offers a brief review of previous empirical literature using the same simulation methodology as we do. Section 4.3 presents the data and discusses the simulation methodology, including the functioning of RTGS systems in general and the specific features of

[1] MNB stands for Magyar Nemzeti Bank, the central bank of Hungary, hereinafter MNB.

VIBER. The indicators used to describe the operation of an RTGS system are explained, along with the assumptions underlying the hypothetical scenarios. The assumptions cover the behavioural reactions of technically non-defaulted participants, the timing and length of operational failures, the number and list of technically defaulted participants and the application of existing back-up procedures. Section 4.4 summarises the simulation results. We simulated altogether six scenarios, three entire-day incidents and three part-time incidents. Section 4.5 presents conclusions and suggestions for further research.

4.1.2 Operational incidents

The smooth functioning of payment and settlement systems assumes the availability of the requisite resources (premises, staff, IT equipment, power, etc.). These resources are exposed to operational risk. They can be endangered by both internal and external factors that could lead to operational disruption. The spectrum of internal and external factors is diverse; it includes power outages, disruptions to telecommunication networks, IT failures (software or hardware problems), natural disasters and terrorist attacks. Resources can be protected by different means (eg back-up sites and procedures or, in the case of outsourcing, by service level agreements). Nevertheless, a considerable amount of money has to be allocated for their protection. Due to the great variety of operational incidents, the unforeseen nature of such events and the low probability of large-scale events, on the level of the entire society, the costs often offset the expected benefits of the protection.

The Hungarian real time gross settlement system fortunately has not yet suffered from any large-scale natural disaster or terrorist attack. The incidents occurring in the recent past had a small impact on the operation of the payment system. The events either affected the components of the central settlement infrastructure or the participant's facilities.

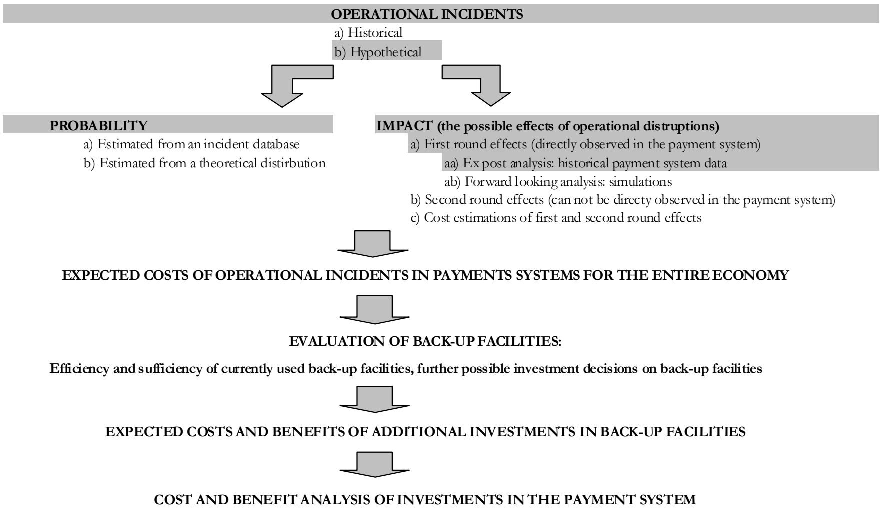
The components of the central settlement system are operated by the central bank and its service providers. Reliable statistics are available on operational incidents affecting these central components. The database of incidents affecting the central components can be considered as complete. Nevertheless, processing of payments also requires the constant ability of VIBER participants to send and receive payment messages. However, the technical problems of VIBER participants are not always reported to the MNB. Exceptions include when the participant asks for prolongation of operating hours.

Despite the incomplete database for external incidents, the MNB is aware that in the recent past several serious operational incidents have occurred. Between April 2003 and May 2007, the MNB recorded 72 external incidents (including problems with monitoring facilities and incidents involving ancillary systems like retail ACH or CSD/SSS), which evidently greatly understates the number of real occurrences. In 20 cases the MNB has information on how long the problems lasted. The average length of time for operational problems was 2 hours and 21 minutes, the maximum 5 hours and 50 minutes, and the minimum 14 minutes. Twenty-six of the 72 cases occurred due to the unavailability of the messaging network, known as SWIFT. For the remainder of the cases, there is no detailed information on the source of the technical problem. Although it is hard to draw conclusions from the incomplete database of the external incidents, it is clear that failures arose at both small and large credit institutions. The incident database shows that failures happen from time to time. Regardless of the low probability of a serious incident, once one occurs, it can have a huge impact on the functioning of VIBER and its participants.

4.1.3 Scope of the research

An operational incident, historical or hypothetical, can be characterised by two dimensions: probability of occurrence and impact (Figure 4.1). The probability of the incident can be estimated from either a long-period incident database or an appropriate theoretical distribution, eg using the extreme value theorem. By measuring the impact of an operational incident, there is a need to distinguish between first and second round effects. First round effects can be directly observed in the payment system. Appropriate indicators of the severity of first round effects include the indicator of payment delay or the volume of unsettled transactions.

Figure 4.1 **Operational incidents and payment system design from risk perspective**



The analysis is ex post if the first round effects are measured by means of historical payment system data, while forward looking if the disruption of the system is forecasted via simulations. Second round effects cannot be directly observed in the payment system. Second round effects materialise in delays of time-critical payments or other contractual obligations. Besides the severity of the incidents, the costs related to those incidents should also be estimated.

After estimating the expected costs of operational incidents in the payment system for the entire economy, one should evaluate the role of the current back-up facilities. Note that there could be other risk mitigating investments, eg limitations or alert systems that might constitute appropriate back-up facilities. The efficiency and sufficiency of the current back-up facilities should be analysed. If the current back-up options prove to be insufficient, investment decisions related to additional back-up options should be considered. The investment decision should rely on a detailed cost-benefit analysis. At the same time, the appropriateness of the payment system design should also be evaluated, and a cost-benefit analysis of the possible risk mitigating investments should be carried out.

In fact, the research carried out can be considered a first step towards evaluating the ability of the payment system to withstand certain types of operational shocks. We set up hypothetical scenarios for operational incidents and examine the severity of the payment system disruption. We do not estimate either the probability of an operational incident or the severity and costs of second round effects. The paper only assesses the severity of first round effects of an operational incident via ex post analysis.

4.2 Literature review

There are several studies prepared by central banks that assess the affects of various operational failures in RTGS or hybrid systems. Bedford et al (2004) assess the impact of different types of operational incidents that could affect the United Kingdom's CHAPS Sterling. The study of Mazars and Woelfel (2005) assesses the impact of a technical default in the French PNS large-value transfer system. Schmitz et al (2006) quantify the contagion effect of an operational incident at one of the participants in ARTIS on the other participants of the system. Bech and Soramäki (2005) evaluate the performance of gridlock resolution algorithms under both normal operating conditions and failure scenarios in the Danish KRONOS system. The study of Enge and Øverli (2006) measures the resilience of the NBO (the Norges Bank's real time large-value settlement system) by varying the liquidity levels.

For analytical purposes simulations based on historical data are carried out. These experiments operate with the payment system simulator developed by the Bank of Finland (BoF-PSS) or Banque de France. With their built-in functionalities these simulators can replicate the functioning of various types of large-value payment systems. The goal of performing the simulations is to shock the payment system and to see what would have happened if the payment system had experienced an unanticipated operational incident. The possibility of insolvency is ruled out. The market participants do not doubt the financial soundness of the institution in trouble; the default is exclusively operational in nature.

All the studies considered address the question whether the inability of a single participant or multiple participants to submit payments leads to serious disturbances in the system. For analysing these effects three features of the incident are given: (1) the set of technically defaulted or affected participants, also named stricken banks; (2) when the incident starts (the timing of the incidents) and (3) how long it takes (the length of the incidents). These parameters are usually set in accord with the worst-case scenarios, given certain constraints (eg the length of the incident should be two hours).

Besides these features, the severity of the failure depends also on the assumptions concerning the liquidity available in the system, the payment system design (eg the existence of back-up procedures) and whether the non-defaulting (or unaffected) participants react to the incident by applying stop sending rules or adjusting bilateral limits.

In general, all studies aim at analysing how operational risk might lead to the realisation of liquidity risk. However, the studies differ in the parameters of stylised incidents, the assumptions made as to system design and behavioural reactions of non-defaulting participants, as well as the indicators of measuring the severity of operational incidents.

4.3 Data and methodology

In this paper the operation of VIBER is analysed with the help of the simulator developed by the Bank of Finland (BoF-PSS). The simulations are conducted under normal (benchmark case or reference scenario) and distressed periods using actual data on payments and liquidity from the period December 2006 – January 2007 (41 business days). The reference scenario replicates the actual functioning of VIBER. The institutional features of VIBER are reflected in the parameterization of the BoF-PSS2.

4.3.1 About RTGS systems with special focus on VIBER

Like CHAPS Sterling, ARTIS, KRONOS or NBO, the Hungarian VIBER is an RTGS system, in which the moments of clearing and settlement are not separated in time; booking is managed item by item continuously and in real time. In VIBER, the processing of payment orders and their final settlement takes place continuously, and the participants concerned are notified in real time. Each settlement takes place by examining whether the participant has provided sufficient liquidity. If so, the payment orders are settled immediately. If the participant does not have sufficient liquidity, the payment will be placed in the central queue. There is one queue per account that contains validated payments waiting for settlement. When entries are placed in the payment queue, they are inserted in FIFO (first in first out) order by priority, the latter being set by the submitter. Payments with high priority are always closer to the front of the queue than those with lower priority. If the payment at the head of the queue cannot be settled due to insufficient cover, the queue is blocked. VIBER has a built-in feature (gridlock resolution algorithm) which solves gridlock situations on a multilateral basis. In VIBER, gridlock resolution can be initiated manually by the central bank, at the request of the submitter. It can also be initiated automatically at a pre-defined frequency. In the current setup gridlock resolution algorithm is initiated every 30 minutes. Basically the algorithm corresponds to multilateral partial offsetting at a given interval of 30 minutes. During the business day, the unsettled payments can be re-prioritised or deleted.

At the end of 2006 VIBER had 38 direct participants. The available (or actual) liquidity of direct participants consists of the current account balance plus the intraday credit line, which can be obtained from the central bank by providing collateral. The list of eligible collateral and the evaluation principles of those assets are determined by the central bank. Assets eligible as collateral for intraday credit are the same as those accepted in monetary policy operations. Pledging additional collateral is possible at any time during VIBER business hours.

In order to replicate the functioning of VIBER the following data were collected:

a) the payments with
   - submission and value dates, time stamps and sequencing parameters that were essential to obtain the submission ranking,
   - the amounts,
   - the original priorities and the changes in priorities,
   - the debited and credited banks, and
   - the message types (customer payments, bank-to-bank items, securities transactions or manual account transfer by the central bank) defined by the submitter,

b) the initial current account balances and intraday credit lines at the opening of the system and changes of the intraday credit lines during the day,

c) the system's opening and closing times, and

d) the timing of the gridlock resolution algorithm.

4.3.2 Specific indicators used for describing the operation of an RTGS system

Besides the usual statistics (turnover and liquidity, concentration indicators, timing indicators) more specific indicators were calculated to describe the operation of VIBER by employing the BoF-PSS.

a) Non-submitted, rejected and unsettled payments

An obvious indicator of the severity of an operational incident is the number and value of payments not submitted to the system. This direct effect is related to the fact that the operational failure of the settlement bank prevents the bank concerned sending payment orders.

The risk resulting from the technical default of a participant can be best captured by the rejected payments indicator. The indicator shows how significant the contagion effect is in terms of number and value of rejected transactions. We refer to unsettled transactions as the sum of the payments not submitted to the system and payments submitted but rejected due to insufficient funds.

b) Hypothetical liquidity levels

Besides the actual level of liquidity, there are various hypothetical liquidity levels which could provide useful insight into the extent to which participants are able to withstand the liquidity risk of an operational incident of a counterparty.

The lower bound of liquidity equals the minimum amount of liquidity required to settle all payments submitted during a day. The lower bound of liquidity corresponds to a very extreme case, in which the banks have just enough liquidity to settle their payments before the end of the day by applying multilateral offsetting as a gridlock resolution algorithm. A more detailed explanation and the formal definition of lower bound can be found in BoF (2005), p. 38.

The upper bound of liquidity corresponds to a liquidity level that a settlement bank would need in order to settle its outgoing payments immediately upon submission. Thus, the upper bound of liquidity is defined as the amount of liquidity needed to settle transactions without any queues.

Potential liquidity is equivalent to the sum of the current account balance and the maximum of the intraday credit line. The maximum of the intraday credit line is based on information concerning eligible assets recorded on the participant's balance sheet. Unfortunately these figures produce only rough estimates, as credit institutions can pledge their securities for other purposes such as to meet collateralization obligations related to stock exchange transactions. Since in Hungary collateral is pledged without transferring the title of the securities, we do not have information on how much of the eligible assets are not available in case additional collateral is needed for payment purposes.

c) Liquidity usage indicator

During the business day the outgoing payments are financed by incoming payments and by the available liquidity, consisting of the current account balance plus the intraday credit line. The liquidity usage indicator (LUI) measures the maximum share of the liquidity that is available for financing payments.[2]

d) Queue and delay statistics

To obtain better insight into the performance of the payment system it is worth analysing the evolution of the queue statistics (BoF, 2005). The percentage of payments settled in real time – in volume and value terms – shows the fluidity of the system. The number and total value of queued transactions show the opposite. Note that in all cases we refer to centrally queued payments. In our analysis we focus on the total value of queued transactions. It should be noted that the indicator is very rough; it does not take into account how large the queue was at certain points in time and for how long the payments were blocked due to insufficient funds.

The maximum queue and the average queue lengths are two indicators that can complement the total value of queued transactions. The maximum queue is the peak queue value during the business day. The average queue length is an indicator that captures the amount of

---

[2] The exact calculation of the liquidity usage indicator (LUI) is as follows. If both the start of day balance (SoDB) and the minimum balance of the day (MB) are negative, than the indicator can be expressed as: LUI = [abs(MB)–abs(SoDB)]/[IDCL+SoDB]. If both the start of day balance (SoDB) and the minimum balance of the day (MB) are positive, than the indicator can be expressed as: LUI = [abs(MB)–abs(SoDB)]/ [IDCL+SoDB], where IDCL stands for intraday credit line. Finally, if the start of day balance (SoDB) is positive and the minimum balance of the day (MB) is negative, than the indicator can be calculated by the formula of LUI = [SoDB+abs(MB)]/(SoDB+IDCL).

time the payments spent in the queue. The average queue length shows the average queue duration of queued payments, namely the total queuing time of payments divided by the total number of queued payments.

The fourth and fairly complex indicator of the queue statistic that was quantified and analysed is the delay indicator. As this indicator has been quantified in many previous studies, it also serves as a good base for international comparison. The delay indicator quantifies the extent to which settlement of individual transactions is actually delayed compared to the theoretic maximum delay to the end of the day. The delay indicator is a relative indicator ranging from 0 to 1. If no transactions are queued, its value is 0. When all transactions are queued to the end of the day, its value is 1. The indicator is calculated as the queuing-time-weighted value of all queued transactions (transaction value multiplied by time in queue) divided by the time-weighted value assuming all payments were delayed to the end of the day. For the formal definition of the delay indicator see BoF (2005), p. 38.

4.3.3 Description of normal functioning of VIBER: the benchmark case

Analysis of the benchmark case, that is the normal functioning of VIBER in December 2006 and January 2007, is valuable for several reasons. First of all, the results of the stress scenarios can be compared with the benchmark case. Secondly, we can identify critical period(s) of the business day(s) when an incident with a pre-defined length might have the largest impact on the functioning of VIBER. Thirdly, we can discover the critical participants based on proxies for liquidity risk and concentration indicators. The critical participants can be either systemically important participants or endangered participants. The technical default of the systemically important participants might have serious negative consequences for the functioning of VIBER. (For a proper definition of systemically important VIBER participants, see Subsection 4.3.3.4.) Endangered participants are institutions which might be heavily influenced by the operational incidents of systemically important participants.

4.3.3.1 Descriptive statistics

Table 4.1 shows the descriptive statistics of VIBER for the months December 2006 and January 2007. All submitted payments were settled by the end of the day; there were no unsettled payments. As demonstrated by Table 4.1, in December 2006 and January 2007 on average 3,429 transactions were settled daily. The number of payments settled ranged from 2,098 to 4,963. The mean value of payments settled totalled 3,496 billion HUF.

Table 4.1 **Descriptive statistics of VIBER (December 2006 – January 2007)**

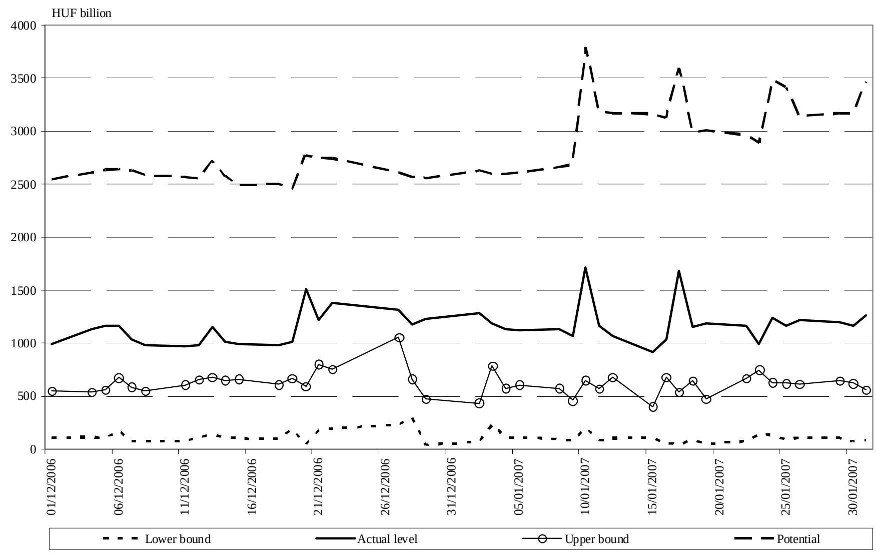
	Minimum	Average	Maximum	St. dev.
Number settled	2,098	3,429	4,963	585
Number unsettled	–	–	–	–
Value settled (million HUF)	1,422,990	3,496,231	5,387,416	751,756
Value unsettled	–	–	–	–

VIBER opens at 8:00 a.m. CET and closes at 5:00 p.m. CET. During the period studied prolongation of operating hours occurred three times (once for 15 minutes and twice for 30 minutes). The credit institutions justified their request for prolongation with reference to technical problems. If we examine the distribution of value of submitted payments according to the system's receipt time, we might conclude that at the beginning of the business day the banks submit relatively more payments. This can be explained by the warehoused payments (payments submitted before the days of the value date) which are channelled into the central accounting system right after the system is opened. By analysing the distribution of the value of submitted payments we can also conclude that the most active submission period for payments initiated by the banks is between 9:30 and 14:00.

4.3.3.2 Liquidity levels

At the aggregate level, the system had more than enough liquidity to settle all payments immediately, and the system as a whole had many securities not posted as collateral. This is reflected in Figure 4.2 by the fact that potential level of liquidity is much higher than the upper bound. In Figure 4.2 it can also be seen that on the system level the actual level of liquidity in VIBER is significantly higher than the lower bound. This means that on the system level the payments can always settle at least by the end of the day. However, there are significant variations across banks. In distressed periods there might be institutions at which the liquidity buffer available might not be sufficient to mitigate the impact of a liquidity shock.

Figure 4.2 **Lower and upper bounds of liquidity, actual liquidity and potential liquidity (billion HUF)**



Note: The data for the Hungarian State Treasury (Single Treasury Account) were included in the liquidity levels.

4.3.3.3 Queue and delay statistics

In December 2006 and January 2007 the total value of queued transactions expressed as a percentage of value settled equalled 16.41% on average (Table 4.2); the maximum was 33.02%. The maximum queue value, expressed as a percentage of value settled, equalled 4.29% on average. However, there was a day when the maximum queue value was 11.08% of all the transactions settled. The mean of the average queue duration of queued payments was 41 minutes. On 19 January 2007, with more than two hours of average queuing time, the indicator reached its maximum.

In relation to the queue statistics there are significant variations across banks. Five participants are responsible for 97% of the queues (both in number and value). Three of the five participants dispose of the highest share in debit turnover. The two participants with the highest proportions of queued payments queued up 76.22% (number) and 70.28% (value) of their payment orders.

The settlement delay was 0.07 on average (Table 4.2), the maximum for the delay indicator equalled 0.16. The settlement delay also varies across banks. There is one small bank whose average settlement delay is remarkably high compared to other system participants. The average settlement delay of this small bank equals 0.1895, which is almost 4 times higher than the average settlement delay of the bank with the second highest settlement delay (0.0478). This can be explained by the specific timing behaviour of the small bank: it sends the majority of its payment orders at the beginning of the day and waits for incoming payments. Meanwhile the payment orders are queued up; the average queue length of the bank is almost 2 hours.

Table 4.2 **Queue and delay statistics (December 2006 – January 2007)**

	Minimum	Average	Maximum
Value of payments initially not submitted	–	–	–
Value of unsettled payments	–	–	–
Total value of queued transactions (in % of value settled)	2.62%	16.41%	33.02%
Maximum queue value (in % of value settled)	1.35%	4.29%	11.08%
Average queue length (hh:mm:ss)	0:08:34	0:41:24	2:08:44
Settlement delay	0.01	0.07	0.16

4.3.3.4 Systemically important and endangered participants

In our analysis those participants are considered as systemically important institutions which dispose of more than 5% of the debit turnover in the reference period. In this way we identified six participants whose technical default might have a serious impact on the functioning of VIBER.[3]

[3] Similarly to our findings, the study of Lublóy (2006) identified the same systemically important institutions according to certain network criteria. In the study a graph theoretical framework was applied. The author argued that the institutions most capable of generating contagion can be best captured by means of valued outdegree centrality and out-proximity centrality. Based on the centrality measures, six institutions were identified; illiquidity of these institutions could cause the most serious disruptions in the payment system.

Participants endangered by the technical default of systemically important members can be identified by assessing the liquidity risk they face under normal conditions and in distressed periods. By examining various proxies for liquidity risk under normal functioning we form clusters with common characteristics. Based on the clustering, we identify those whose members could be easily endangered in the case of a liquidity shock. As a next step, the circle of these endangered participants is enlarged by means of a simple sensitivity analysis based on a stress indicator. As a result, we arrive at an indicative list of participants who might be endangered if a systemically important member technically defaults.

Based on the queue, delay and liquidity indicators, we formed five clusters. Our results are summarised in Table 4.3. The first column of the table states the share of days during the analysing period (41 days) when the actual level of liquidity was less than the upper bound of liquidity. When the liquidity level is at or above the upper bound, the settlement of transactions can happen without any queues. The higher the number of days on which the actual level of liquidity was less than the upper bound of liquidity, the higher was the probability of having longer queues and delays in the bank's payment process and hence the lower was the level of the bank's liquidity buffer. This also means that if the bank looses some of its credited payments, it could lead to more unsettled payments. The second column of Table 4.3 provides additional criteria for dividing the VIBER participants into smaller subgroups. The grouping criteria in the second column are not homogeneous in the case of the three main groups obtained according to the criteria of the first column. The criteria rather reflect points in a decision tree for forming subgroups of homogeneous banks within the main group.

Participants in Group 'A' have a higher actual level of liquidity than the upper bound of liquidity on most days. Participants in Group 'A' have recorded queues extremely rarely. The participants are not prone to liquidity risk, the actual liquidity level already ensures the existence of a buffer against liquidity risk. Additionally, the participants have high liquidity buffers in their balance sheets, in case the actual liquidity proves to be insufficient.

Table 4.3 **Grouping of VIBER participants**

<table>
<tr><th colspan="3">Share of days when actual liquidity was below the upper bound of liquidity</th><th colspan="2" rowspan="2">Additional criteria applied for grouping VIBER participants</th><th rowspan="2">Number of participants</th><th rowspan="2">Share in debit turnover</th><th rowspan="2">Types of institutions in the group</th><th rowspan="2">Groups</th></tr>
<tr><th colspan="3">(in % of 41 days</th></tr>
<tr><td rowspan="4">0%</td><td rowspan="4"><...<</td><td rowspan="4">10%</td><td colspan="2">Liquidity usage indicator greater than 50% on at least half of the days</td><td>5</td><td>13.51%</td><td>Banks (sometimes with sepcial profile)</td><td rowspan="4">A</td></tr>
<tr><td rowspan="3">Liquidity usage indicator less than 50% on at least half of the days</td><td>Ratio of intraday credit line to available liquidity is low</td><td>8</td><td rowspan="3">21.14%</td><td>Universal and specialized banks, Hungarian State Treasury</td></tr>
<tr><td>Ratio of intraday credit line to available liquidity is mid-sized</td><td>8</td><td>Banks (sometimes with special profile)</td></tr>
<tr><td>Ratio of intraday credit line to available liquidity is high</td><td>4</td><td>Specialized credit institutions</td></tr>
<tr><td rowspan="2">10%</td><td rowspan="2"><...<</td><td rowspan="2">50%</td><td colspan="2">Balance sheet seems to contain enough liquidity buffer</td><td>3</td><td>0.25%</td><td>Banks (often with special profile)</td><td>B</td></tr>
<tr><td colspan="2">Rarely are there days when balance sheet does not have enough liquidity buffer</td><td>3</td><td>3.74%</td><td>Banks (sometimes with special profile)</td><td>C</td></tr>
<tr><td rowspan="6">50%</td><td rowspan="6"><...<</td><td rowspan="6">100%</td><td rowspan="3">Balance sheet seems to contain liquidity buffer, but there may be days when it is not sufficient</td><td>Delay indicatoris relatively low</td><td>0</td><td rowspan="6">55.43%</td><td rowspan="6">Foreign owned banks highly exposed to FX settlement risk</td><td rowspan="3">D</td></tr>
<tr><td>Delay indicatoris mid-sized</td><td>3</td></tr>
<tr><td>Delay indicatoris relatively high</td><td>0</td></tr>
<tr><td rowspan="3">It can easily happen that balance sheet does not have enough liquidity</td><td>Delay indicatoris relatively low</td><td>0</td><td rowspan="3">E</td></tr>
<tr><td>Delay indicatoris mid-sized</td><td>1</td></tr>
<tr><td>Delay indicatoris relatively high</td><td>1</td></tr>
</table>

Please note that two participants (central bank and Hungarian Post) are not clustered.

Participants in Group ‘B’ queue up transactions more often, but their balance sheet contains abundant liquidity buffer. In the reference period these institutions did not post any collateral. Based on debit turnover, these members are the least active participants.

Group ‘C’ contains those participants which queue up payments more often than participants in Group ‘A’. In the case of liquidity shocks there might be occasions when potential liquidity is not enough for immediate settlement of payments. We consider these credit institutions as endangered participants, as they might face problems if there are liquidity shocks.

Participants in Group ‘D’ build up queues often and transactions spend quite some time in the queue. The delay indicator is medium-sized for them. Their balance sheet contains some liquidity buffer. Taking the uncertainty regarding the proper level of potential liquidity into account, these institutions might easily experience problems if there is a liquidity shock. We consider these credit institutions as endangered participants.

The two participants in Group 'E' have no liquidity buffer in their balance sheets. Participants in this group dispose of the highest delay indicator. One of the institutions has a significantly greater indicator. However, its share in debit turnover is low. The other credit institution is the most active one in VIBER (in monthly average its share of credited and debited items is the highest in value terms). This institution is certainly pushed toward more active payment and liquidity management. This participant manages its intraday credit line actively.

Using a simple sensitivity analysis, we examined what happens to VIBER participants if they loose a pre-defined proportion of their incoming payments. We measured whether the participants' eligible asset portfolio would be sufficient for financing incoming funds. We found that eleven institutions would face serious problems on at least 25% of the business days if they did not get at least 50% of their incoming payments. These participants rely heavily on financing via incoming payments. Interestingly, during the clustering we classified 8 of these 11 as endangered participants. In line with this simple sensitivity analysis, we enlarged the group of endangered participants by the three previously unidentified institutions.

If we look at the lists of systemically important and endangered institutions, we notice an overlap between them. Only one of the systemically important participants is not identified as an endangered institution.

In sum, the six most active players measured by debit and credit turnover in VIBER could easily trigger a serious liquidity shock. On the other hand, these institutions could easily suffer from a liquidity shock if that is caused by a systemically important member. The six most active institutions, and in many cases their important customers (eg large foreign financial institutions for whom domestic banks are HUF correspondents), are active money market participants in the FX swap segment. A recent FX settlement survey by the central bank (Tanai, 2007) showed that these participants recorded the largest FX settlement risk in the domestic banking sector. These credit institutions initiate large amount of transactions in VIBER with a relatively small balance sheet. Should a liquidity shock arise in the system (eg due to improper functioning of financial markets or operational incidents), their balance sheets could easily be a bottleneck. On the one hand, the relatively high velocity of liquidity makes the institutions vulnerable to unexpected liquidity shocks but also indicates extremely efficient payment and liquidity management practices.

4.3.4 Assumptions of the different stress scenarios

In the paper operational failure is defined as the existence of technical problems for one or more system participants other than the MNB. The technically defaulted participants are unable to send and receive payment messages, ie they are unable to access the SWIFT network. Inability of technically defaulted participants to send and receive payment messages would not hinder other non-defaulted participants from sending payments to the problematic credit institutions.

The liquidity risk caused by operational disruptions is examined via several scenarios. The scenarios are hypothetical and based on stylised operational failures, though they might not be far from reality. The scenarios are based on different assumptions concerning:

- the behavioural reactions of technically non-defaulted participants,
- the timing and length of the operational failures,
- the number of technically defaulted, systemically important participants, and
- application of existing back-up procedures.

4.3.4.1 Behavioural reactions of technically non-defaulted participants

In the simulations several behavioural assumptions are made. On the one hand, it is assumed that the behaviour and payment patterns of the non-defaulted VIBER participants are not affected by the initial shock. This behavioural assumption of no reaction has several dimensions. Firstly, it is assumed that the settlement banks wish to settle the same volume and value of transactions (even to the technically defaulted participant) with the same priority as under normal business conditions. Secondly, in the model, the settlement banks do not raise additional liquidity by borrowing funds from the parent bank or the interbank market in order to ensure settlement of all their payment orders. Based on inquiry-responses from some Hungarian credit institutions, this behavioural assumption is not far from reality for (already) agreed obligations. There might be several reasons why banks do not stop sending payments to the stricken bank. Firstly, the technical infrastructure of the payment systems located in the back offices is very complex. Most of the payment orders are generated automatically from several internal back office systems as the last step of the straight through processing (STP). As a consequence, it is not easy to modify the list of payment orders waiting for transmission to the central settlement engine. Secondly, the banks cannot take risks with their reputations. If a bank wishes to maintain its high prestige, it should prove that it is able to fulfil its contractual commitments, even in the face of problems. Thirdly, a contractual commitment is an obligation which, if not met, could have legal and financial consequences.

On the other hand, in separate scenarios we assume that the settlement banks are not passive economic agents. Instead they take actions to prevent the bank under distress from becoming a liquidity sink and they stop sending payments to the bank experiencing the technical default (stop sending rule). In real life it usually takes time to get information about participants' technical problems. Anecdotal evidence suggests that there is a time-lag between a VIBER participant experiencing an operational failure and the timing of phone calls in which the technically defaulted bank announces its problems to the MNB. It can also happen that, even before the announcement, the intact settlement banks notice that they are not receiving expected incoming payments and want to know whether the MNB has any information about the source of the problem at the settlement bank under distress. They usually call the central bank in order to obtain more information. Empirical evidence suggests that, even if the intact members become fully informed, they usually do not know whether they should stop sending payments. The central bank usually encourages them to submit their payments. There is high degree of uncertainty about how much time the elimination of information asymmetry requires. In our simulations we assume two hours reaction time, thus the stop sending rule applied in two hours is simulated. To examine scenarios with more efficient alarming systems would not be feasible.

4.3.4.2 The timing and the length of operational failures

The timing and length of operational incidents have a large impact on system performance. In one set of simulations the worst-case scenarios are considered. It is assumed that the operational incident occurs no later than the opening of VIBER and lasts until the end of the business day (entire day incident). Thus, the technical problem arose no later than 8:00 a.m. CET and could not be sorted out until 5:00 p.m. CET.

We also ran simulations where the starting point of the operational failure is not fixed in advance but the length of the incident is set at a pre-defined level, 4 or six hours respectively (part-time incidents). In this case the timing of the operational incident is the outcome of a worst-case scenario maximization routine. We look for an appropriate algorithm to find the worst-case scenario; the operational failure should arise when it has the largest negative impact on the system.

However, there are some constraints. The disruption of payment processing can be sufficiently severe only if the banks wish to settle many payments during the rest of the business day. As a consequence, it is of no interest to analyse the impact of an incident that happens close to the end of the business day, as the value of remaining transactions to be settled would be low. The daily cumulative distribution of payment flows showed that in VIBER most payments (80–90% of turnover) are settled by 2 p.m. CET. Thus, it is not useful to let the operational incident occur later than that.

4.3.4.3 The number and list of technically defaulted participants

Initially it was assumed that a single settlement bank is unable to submit payments to VIBER owing to a failure of its internal back-office system. The technical problem is isolated; a single bank is hit by the shock. As we are interested in the disturbance of the payment system in the worst case, there is no point in choosing a minor participant whose failure would not have a significant impact on the functioning of VIBER. Thus, we assume that systemically important institutions are affected by the technical problem.[4] In this paper, as we focus on the worst case, we quantify the impact of the first six systemically most important institutions.

Via additional scenarios we also quantify the risk implications of an operational disruption affecting the ability of multiple settlement banks to submit payments to VIBER. If we selected two banks randomly from the 38 participants, we would end up with 703 possible combinations. If we selected three banks randomly we would arrive at

[4] Note that occasionally it might happen that operational problems at a bank with a relatively lower debit turnover results in a severe disruption of the payment system. The connectedness of the bank and the unequal distribution of liquidity might also play a crucial role. Nevertheless, the probability of a small bank generating a large shock is relatively low.

8,436 various combinations. The modification of the input data and, in the case of a part-time incident, the worst-case scenario maximisation procedure would take too long for all of these combinations. As a result, we opted for analysing the worst cases and we assumed that the technical problems hit two of the six systemically most important institutions. This resulted in 15 possible pairs of banks suffering from a technical problem simultaneously. It is important to stress that the probability of such an incident is low.

4.3.4.4 Possible contingency procedures

The functioning of the system in a distressed situation can be enhanced if back-up procedures are in place. In some countries these contingency procedures usually require time-consuming manual intervention. There exist quicker (mostly electronic) back-up facilities, but this is not the case in Hungary. We assume that one or more participants cannot submit their payments via SWIFT, but communicating via fax, which is the back-up for SWIFT, is feasible. The identification of not-yet-sent payments, the production of the paper-based credit transfer with all the required data and the submission of the fax to the central bank requires some time, like the processing of these fax-based transactions at the central bank.[5] It is important to note that, due to the fear of duplicating payments, participants hesitate to use back-up options.

4.4 Simulation results

4.4.1 General considerations on simulation techniques

The technical default of one or more VIBER participants has both direct and indirect effects on system performance. The direct effect is obvious: operational failure of the settlement bank prevents the concerned bank from submitting payments to the system. If the problem is severe enough and cannot be fixed before the end of the business day, in the absence of contingency arrangements, the

[5] The work at the central bank takes approximately five minutes per transaction. During the simulations we assumed four persons as an average number of staff members in the VIBER team of the central bank. This means that in one hour the VIBER team is able to process approximately 50 fax-based transactions.

settlement bank under distress will end up with many unsettled transactions. In addition to this direct effect, there might be significant indirect, so-called network effects. Namely, as the liquidity position of each VIBER participant is influenced by payment flows, if one settlement bank is unable to send payments, the liquidity position of the remaining participants is threatened. This, in turn, could force the initially unaffected settlement banks to queue and thus delay payments. It might also happen that a large amount of liquidity is drained out of the system, as it is accumulated on the account of the bank experiencing the technical problem. As a consequence, there might be banks short of liquidity, and they might end up with rejected payments at the end of the day. Both direct and indirect impacts are captured by the performance indicators.

For simulation purposes, on the day of the incident, the transactions initiated by the technically defaulted participants should be removed or modified. The transactions should be removed if it is assumed that the operational problem cannot be solved before the end of the day. The transactions should be modified if it is assumed that the bank managed to solve the operational problem within the business day. In this latter case the time stamps (receipt times) of the transactions should be altered. The modified time stamps should reflect the point in time when the bank managed to sort out the technical problem.

Payment instructions initiated by technically functioning participants (including infrastructure like the CSD/SSS) to be debited or credited on the account of the defaulted participants are processed normally. Transactions processed normally include mandated payments.[6] In addition, warehoused outgoing payments of technically defaulted participant are processed normally as well.

If it is assumed that the initially unaffected participants do not react to the shock, no modification in the initial dataset is needed. However, the participants may react to the liquidity shock. If the banks apply the stop sending rule, the time stamp of the transactions which are sent two hours after the beginning of the incident should be

[6] At default, the transactions are initiated by the debited bank. However, there are cases when payments are initiated by someone else. These payments are called mandated payments and are initiated, for instance, by the central bank (eg settlement of clearing positions of the ancillary systems) or by the CSD/SSS (eg DVP transactions). Mandated payments include the START transfers, the DVP transactions initiated by KELER, the multinet settlement of stock exchange deals, the settlement of card transactions and cash withdrawals with the central bank, etc.

modified. The new time stamp of the transactions should reflect the time of the restoration of the SWIFT connection.

In the simulations it is implicitly assumed that the banks are able to solve the problem right after the official closing of VIBER and somehow the payments are settled. We do not simulate two- or three-day operational incidents.

4.4.2 Simulated scenarios

The simulations of operational failure scenarios are based on a combination of assumptions (see Subection 4.3.4). The assumptions applied in the simulations are summarised in Table 4.4. In the first simulation setup (Scenario 1 to 3) entire-day incidents were imitated. It was assumed that the operational failure starts at the beginning of the day and the bank under distress cannot sort out the problem until the end of the business day. In each scenario one of the six systemically most important banks became unable to send payment orders. In Scenario 1, shown in the first column of Table 4.4, no back-up facilities and behavioural reactions were assumed. In Scenario 2 we examine the shock-mitigating impact of back-up facilities. In Scenario 3 the disturbance of the payments system was assessed if technically non-defaulted participants took actions to prevent the bank under distress from becoming a liquidity sink, and stopped sending payments to the bank experiencing the technical default after two hours (stop sending rule).

Table 4.4 **Scenarios examined in simulation exercises**

	Entire day incident			Part-time incidents		
Scenario	1	2	3	4	5	6
Number of technically defaulted participants	1	1	1	1	1	2
Duration of the incident (hours)	9	9	9	4	6	4
Contingency procedures: Back-up facilities	–	+	–	–	–	–
Behavioral reaction of technically non-defaulted participants	–	–	+	–	–	–

In the second simulation setup more realistic part-time incidents were simulated. We looked for worst-case scenarios in which operational failures of given length occurred when the value weighted submission delay for payment orders sent by one of the six systemically most important banks was the highest. In the worst-case scenario maximisation procedure, the values of postponed payments were weighted by the time lag between submission and the end of the operational incident. The weighing of payments by the time left until the end of the incident takes into account both the value of the payment and the delay in submission (consequently the delay in settlement).

In the worst-case scenario maximisation procedure, only the length of the incident and the number of technically-defaulted banks were provided as input parameters. It was assumed that the technical problem can be sorted out in either four or in six hours, but no later than the end of the business day.[7] In Scenarios 4 and 5 one bank defaulted technically and the incident lasted for four or six hours respectively. Scenario 6 corresponds to a situation where an operational problem affected two banks simultaneously for four hours. It was assumed that the incidents began at the same moment. For the sake of simplicity, it was also assumed that the technical problems were sorted out within the same time period.

The output of the worst-case scenario maximisation gave us the starting point for the operational failure and the list of technically defaulted participants. The timing and the technically defaulted bank(s) were not necessarily the same across the days. They were highly dependent on the daily payments patterns.

4.4.3 Disturbance in the payment system: entire-day incidents (Scenarios 1–3)

4.4.3.1 Scenario 1: Entire-day incidents – no back-up facilities and no behavioural reactions

The simulation results for the first scenario are summarised in Tables 4.5 and 4.6. As a first step it was assumed that the bank with the highest turnover suffers from an operational incident and is unable to submit payments during the entire business day. The counterparties did not change their behaviour and back-up facilities were not in place. Detailed results of the simulations accessing the impact of the

[7] Quantifying the effects of incidents lasting for three or five hours, for example, could also have been reasonable and realistic. However, in order to reduce the number of simulated scenarios, we disregarded these possibilities. Further research could be done in this direction as well.

other five technically defaulted participants on the payment system are presented in Appendix 1 of Lublóy and Tanai (2008).

As shown in Table 4.5, due to the technical default of Bank 1 an average 16.3% of the payments could not be submitted to the system. On 6 of the 41 days more than 20% of the payments sent in the benchmark scenario were not submitted at all. Meanwhile, there were no rejected payments in the benchmark scenario. In this scenario on average 16.2% of the submitted payments remained unsettled. As shown in the last column of Table 4.5, in the very extreme worst-case scenario, 21.7% of the payments were not submitted and 34.7% of the payments submitted were rejected. In comparison with the benchmark scenario, on average altogether 31.0% of the payments remained unsettled, while in the worst case 50.9% of the payments remained unsettled (either not submitted or rejected). The indicator of multiplication equals the ratio of the value of rejected payments over the value of non-submitted payments. The indicator reflects the following: one unit of non-submitted payment resulted in x units of rejected payment. In the worst-case the multiplication effect equalled 1.25, meaning that a single unit of non-submitted payment resulted in 1.25 units of rejected payment.

In an international comparison,[8] the proportion of unsettled transactions is remarkably high. This can be explained by the high concentration of debit and credit turnover and by high rank correlations. The banks – being relatively poorly endowed with liquidity – rely consciously on financing via incoming payments. Thus, if there is a deficit in the incoming payments, outgoing payments cannot be settled due to lack of funds. Not surprisingly, the five banks disposing of the highest value of unsettled transactions include the banks with the highest debit and credit turnover in VIBER. They account for around 88% of unsettled payments. Due to the liquidity drain effect, these banks are left with 26% of rejected payments on average at the end of each day. Four other small banks, having close relations with the banks active in the payment system, suffer heavily from the liquidity drain effect as well. They cannot send 16% to 34% of their payment orders respectively. All of the nine banks can be found in the list of possibly endangered participants.

Another important explanation of the high proportion of unsettled transactions might be linked to the structure and size of the money markets. In the euro zone several large market participants (eg JP

[8] See eg the studies of Bedford et al (2004), Mazars and Woefel (2005), Schmitz et al (2006), Ledrut (2007) or the Financial Stability Review of the National Bank of Belgium (2007).

Morgan, Morgan Stanley) might be considered 'small', at least compared to the size of the entire euro market. In contrast, in Hungary the above-mentioned large market participants – involved in some HUF deals through their correspondent banks – might be considered large, especially relative to the size of the market.

Table 4.5 **Disturbance in payment system: operational incident at Bank 1**

Bank 1 – Entire day incident	Minimum	Average	Maximum
Value of non-submitted payments (in % of the benchmark scenario)	4.6%	16.3%	21.7%
Value of rejected payments (in % of submitted payments)	0%	16.2%	34.7%
Value of unsettled payments (in % of the benchmark scenario)	2.0%	31.0%	50.9%
Multiplication effect	0.00	0.83	1.25
Total value of queued transactions	3.4%	38.4%	54.0%
(in % of submitted payments)	(1.29)	(2.34)	(1.64)
Maximum queue value	2.7%	19.4%	42.0%
(in % of submitted payments)	(2.00)	(4.52)	(3.79)
Average queue length (hh:mm:ss)	0:55:12	1:49:41	2:35:21
	(6.44)	(2.65)	(1:21)
Settlement delay	0.13	0.29	0.5
	(13.00)	(4.42)	(3.13)

Compared to the benchmark case, the total value of queued transactions, measured as the percentage of submitted payments, was approximately 2.34 times higher. Note that in the last four rows the figures in the brackets show how many times the indicator became larger in the case of the disturbance compared to the normal operation of VIBER. In the worst case, more than half (54.0%) of the submitted payments were in the queue once, for either a shorter or longer period. Both the average and the maximum of the maximum queue value indicators increased significantly, to 4.52 and 3.79 times higher respectively. The average queue length increased drastically as well, to 2.65 times higher than average. If Bank 1 suffered from an operational incident, the average of the settlement delay indicator equalled 0.29. This is more than 4 times higher than in the benchmark case. Note that the minimum of the delay indicator is 13 times higher than in the benchmark case.

Table 4.6 compares the outcomes of the simulations where one of the six systemically important participants suffers from an operational incident. The values shown in Table 4.6 are averaged over 41 days. The minimum and the maximum of the corresponding indicators are shown in Appendix 1 of Lublóy and Tanai (2008). Evidently, the lower the turnover of the participants in VIBER, the lower the value of payments not submitted to the system (measured as a percentage of value of payments of the benchmark scenario). The value of unsettled payments also decreases slightly. In the case of the bank with the highest turnover, on average 16.2% of the submitted payments were rejected, while in the remainder of the cases the corresponding figure ranges from 0.5% to 13.8%. The multiplication effect is strongest in the case of the technical default of Bank 2, on average 1 unit if non-submitted payments generated more than one unit of rejected payment.

Table 4.6 **Disturbance of payment system in Scenario 1**

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Value of non-submitted payments (in % of the benchmark scenario)	16.3%	13.7%	10.3%	6.6%	5.8%	4.5%
Value of rejected payments (in % of submitted payments)	16.2%	13.8%	7.0%	3.1%	2.7%	0.5%
Value of unsettled payments (in % of the benchmark scenario)	31.0%	26.7%	17.2%	9.9%	8.6%	5.1%
Multiplication effect	0.83	1.01	0.68	0.47	0.46	0.11
Total value of queued transactions	38.4%	39.4%	34.2%	30.5%	25.4%	22.8%
(in % of submitted payments)	(2.34)	(2.40)	(2.08)	(1.86)	(1.55)	(1.39)
Maximu queue value	19.4%	17.8%	13.08%	9.4%	7.6%	6.1%
(in % of submitted payments)	(4.52)	(4.14)	(3.05)	(2.20)	(1.78)	(1.42)
Average queue length (hh:mm:ss)	1:49:41	2:07:23	1:27:39	1:07:35	0:59:17	0:43:51
	(2.65)	(3.08)	(2.12)	(1.63)	(1.43)	(1.06)
Settlement delay	0.29	0.27	0.20	0.12	0.10	0.08
	(4.42)	(4.06)	(2.97)	(1.76)	(1.54)	(1.21)

The total value of queued transactions is almost the same for the first three banks. If an operational incident hits one of the three banks with the largest turnover in VIBER, one third of the transactions are queued up for some time during the business day. The corresponding figure is one fifth for the bank with the sixth largest turnover in VIBER. The maximum queue value, the average queue length and the settlement also decrease significantly in relation to turnover. In general, queue and delay indicators show a more favourable picture with a lesser role for the shocked VIBER participants.

The message from the simulation exercise is straightforward. Regarding liquidity levels and unchanged timing behaviour, an operational incident at the most active players can lead to serious disturbances in VIBER. The disturbance highly depends on the daily payment patterns; the severity caused by the incident changes from day to day. Operational incidents at the three systemically most important banks require special attention, especially if their problem cannot be sorted out until the end of the day. There are several lines of defence which can mitigate the severity of such operational incidents. The first includes (electronic) back-up facilities. If they work properly, the incident might not even be noticed by other VIBER participants. In the next subsection we examine the impact of the time-consuming, paper-based back-up solution. The second factor which can provide some protection for the intact members lies in adapting to the situation. It is important to note that the technically non-defaulted banks would probably trade with operationally viable counterparties more than with the bank in a distressed situation. If this is the case, the payment pattern is changed as part of the intraday trade is adjusted. Since the intraday financial market trades were not identified in the paper, we overestimated the consequences of operational incidents.

4.4.3.2 Scenario 2: Entire-day incidents – Back-up facilities without behavioural reactions

In this section we examine the shock-mitigating impact of back-up facilities. We assume that effective business continuity arrangements are in place and back-up options are employed one hour before system closing. The processing of payments was assumed to be carried out manually. Thus, we do not take account of the possibility that the back-up options enable settlement of a very large number of payments before system closing. In the simulations it was assumed that the banks under distress can submit altogether 50 payments. The payments were ranked initially by priorities set by the banks, secondly by the amount of the transactions and thirdly by the receipt time of the central accounting system. The first 50 transactions in this ranking were submitted to VIBER one hour before closing of the system. Note that implicitly it was assumed that the internal systems of the stricken banks work properly and thus the stricken banks have up-to-date information about their payment obligations.

Table 4.7 summarises the disturbance in the payment system under Scenario 2. The values shown in Table 4.7 are averaged over 41 days. The minimum and the maximum values of the corresponding indicators are shown in Appendix 2 of Lublóy and Tanai (2008). If Bank 1 submits 50 of its daily 482 payments at the end of the business day, then only 3.0% of the payments were not submitted to the system instead of the 16.3% of the payments in Scenario 1. (For comparison see Table 4.6.) The value of rejected payments also decreased significantly, from 16.2% to 0.1%. There is no such significant decline in the value of the indicators relating to the queue statistics. The significant queues and delays in the system can be explained by the fact that the payments of the technically defaulted bank are channelled to the system in the last hour of the business day.

Table 4.7 **Disturbance of payment system in Scenario 2**

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Value of non-submitted payments (in % of the benchmark scenario)	3.0%	8.2%	5.0%	3.1%	0.2%	0.1%
Value of rejected payments (in % of submitted payments)	0.1%	5.8%	2.3%	1.0%	0%	0%
Value of unsettled payments (in % of the benchmark scenario)	3.3%	14.5%	7.5%	4.4%	0.3%	0.1%
Multiplication effect	0.03	0.71	0.45	0.31	0.00	0.04
Total value of queued transactions	32.1%	36.7%	31.9%	29.5%	23.0%	21.6%
(in % of submitted payments)	(1.96)	(2.23)	(1.95)	(1.80)	(1.4)	(1.32)
Maximu queue value	16.4%	16.5%	12.3%	9.1%	7.0%	5.8%
(in % of submitted payments)	(3.83)	(3.84)	(2.87)	(2.11)	(1.63)	(1.36)
Average queue length (hh:mm:ss)	1:20:28	1:08:39	1:14:33	1:14:33	0:52:42	0:42:28
	(1.94)	(1.66)	(1.80)	(1.80)	(1.27)	(1.03)
Settlement delay	0.24	0.18	0.18	0.11	0.09	0.08
	(3.66)	(2.76)	(2.69)	(1.67)	(1.39)	(1.17)

If Bank 2 or Bank 3 suffers from the operational incident but employs back-up facilities, then the decline in the proportion of unsubmitted and rejected payments is significantly lower than in the case of Bank 1. This can be explained by the fact that the 50 transactions that were booked manually do not have the highest values. On several days Bank 2 and Bank 3 had many customer payments with high priority. As the values of these payments are much lower than the values of bank-to-bank payments, it could be that the payments with the far highest values were not submitted to the system. On these days, the liquidity drain effect showed a very similar pattern to that of Scenario 1.

As shown in Table 4.7, if Bank 4 is under distress but uses the back-up facility, then the disturbance to the payments system is somehow similar to the disturbance caused by Bank 1. If Bank 5 and Bank 6 are technically defaulted, but 50 of their transactions are booked manually at the end of the business day, there will be almost no rejected payments. In contrast, compared to the benchmark scenario, the queues became larger and lasted longer. This is reflected in the figures in brackets, which tell how many times the indicator became larger in a distressed situation compared to the normal operation of VIBER.

If we compare the results of the entire-day incident with and without back-up facilities, we might conclude that even if the number of payments processed manually is limited, the effect of submission to the system of the payments stuck in the internal queues of the technically defaulted banks is positive. This is obvious if we compare Table 4.6 with Table 4.7. Nevertheless, the improvement depends heavily on the selection procedure of manually processed payments. We have chosen a prioritization scheme that leads to significant improvement if the incident happens at Bank 1, Bank 5 or Bank 6. The disruption of the payment system remained almost unchanged when operational failure occurred at Bank 2, Bank 3 or Bank 4. In sum, the improvement achieved by back-up facilities is highly dependent on how the technically defaulted institution chooses the payments to be processed manually in a distressed situation.

4.4.3.3 Scenario 3: Entire-day incidents – Behavioural reactions without back-up facilities

Table 4.8 shows the disturbance of the payments system if one of the banks is hit by an operational incident and the rest of the system reacts by blocking their payments to the stricken bank within two hours. Comparing Table 4.8 with Table 4.6, we conclude that by applying the stop sending rule the non-submitted payments increased significantly and the value of rejected payments decreased drastically. This fact is also nicely captured by the indicator of multiplication that remained below 0.1 in all cases. The composition of unsettled payments is very different. In Scenario 1 a high proportion of payments is rejected due to insufficient liquidity. The liquidity drain effect is significant; many banks suffer from the impact of the technical default. In contrast, in Scenario 3 banks try to escape from the liquidity drain effect by not submitting their payments to the bank under distress. Many payments are withheld; the high disruption of the payment system can be explained by this behaviour. It is important to stress that, despite the higher proportion of unsettled payments in Scenario 3, the shock remains isolated. The disruption of the payment system is strongly connected to payments not submitted by the bank under distress and to payments not submitted to the bank under distress.

The value of payments that remained unsettled (unsubmitted or rejected) is higher in Scenario 3 than in Scenario 1. This is the price of isolating the shock and privileging payments sent to other participants.

The higher value of unsettled payments means that some kept-back payments could have been settled. Comparing the evolution of queues and delays in Scenario 1 and Scenario 3, we concluded that in Scenario 3 the transactions are settled more smoothly and the queues and the delays are shorter. For more detailed insight into the disturbance of the payment system under Scenario 3, see Appendix 3 of Lublóy and Tanai (2008).

Table 4.8 **Disturbance of payment system in Scenario 3**

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Value of non-submitted payments (in % of the benchmark scenario)	29.6%	25.5%	18.4%	12.1%	10.9%	8.0%
Value of rejected payments (in % of submitted payments)	2.6%	2.0%	1.6%	0.7%	0.5%	0.2%
Value of unsettled payments (in % of the benchmark scenario)	32.5%	28.2%	20.4%	3.3%	11.8%	8.5%
Multiplication effect	0.09	0.08	0.09	0.06	0.04	0.03
Total value of queued transactions	19.5%	23.5%	22.8%	20.8%	17.1%	18.0%
(in % of submitted payments)	(1.19)	(1.43)	(1.39)	(1.27)	(1.04)	(1.10)
Maximu queue value	7.1%	7.8%	7.5%	5.9%	4.8%	4.9%
(in % of submitted payments)	(1.64)	(1.80)	(1.74)	(1.37)	(1.11)	(1.15)
Average queue length (hh:mm:ss)	0:51:35	1:13:18	1:17:31	0:46:01	0:47:18	0:45:23
	(1.25)	(1.77)	(1.87)	(1.11)	(1.14)	(1.10)
Settlement delay	0.11	0.13	0.11	0.08	0.07	0.07
	(1.67)	(1.99)	(1.72)	(1.18)	(1.12)	(1.13)

As Scenario 3 relies on behavioural expectations, the results should be analysed with caution. We applied stop sending rules in the simulations without filtering out the intraday financial transactions (transactions agreed and settled on the day of the incident), which could be misleading. In order to see the influence of the stop sending rule, intraday financial transactions should be mapped. In addition, we think that stop sending is very doubtful behaviour. It is more realistic to assume that operationally viable participants fulfil their agreed obligations in the course of the day and payments sent to the technically defaulted bank are placed to the end of the internal or external queues. It could also happen that transactions in question are only submitted at the end of the business day. If in the absence incoming payments the participants face liquidity deficits and cannot finance all of their outgoing transactions, they would try to finance payments to the participant under distress from payments to be received from the participant suffering from the shock. Consequently, management of the loss reallocation would be easier afterwards. Usually the loss reallocation works very smoothly between participants. Further research is required to map changes in intraday trading patterns and to assess how liquidity managers adapt to the situation.

4.4.4 Disturbance in the payment system: part-time incidents (Scenarios 4–6)

In Scenarios 4 and 5 it was assumed that one bank suffers from an operational incident and is unable to send payment orders to its counterparties for four and six hours respectively. The beginning and the end of the four- and six-hour intervals were calculated via the worst-case scenario maximisation procedure detailed in Subsection 4.4.2. Of the 41 days analysed, in Scenarios 4 and 5, on 35 days and 32 days respectively the bank most active in the payment system would have to suffer from on operational incident in order to cause the most severe disturbance in the payment system. On the other the days, the second, third and fourth participants most active in the payment system would have to suffer from the operational incident in order to have the most serious impact on the payment system. In Scenario 6, in most of the cases, the joint technical default of Banks 1 and 2 (20 cases) vs. Banks 1 and 3 (17 cases) generated the highest values of payments not submitted on time.

Table 4.9 shows the value of transactions not submitted on time in Scenarios 4 to 6. The maximum, the average and the minimum of the value of payments with delayed submission are displayed. Obviously, if the incident lasts longer or two banks default technically, the value of transactions not submitted on time increases. In Scenarios 4, 5 and 6, the average value of transactions not submitted on time equals 466 billion, 541 billion and 806 billion HUF respectively.

Table 4.9 **Value of transactions not submitted on time (million HUF)**

	Minimum	Average	Maximum
Scenario 4	75 040	466 334	650 003
Scenario 5	98 459	540 828	773 256
Scenario 6	124 417	806 287	1 186 135

Table 4.10 provides some information on the timing of incidents. The beginnings of the operational incidents occurring the earliest and latest on average are highlighted. If the incident lasts six hours, it should start earlier compared to an incident lasting for four hours.

Table 4.10 **Timing of incidents**

	Minimum	Average	Maximum
Scenario 4	8:52:16	9:54:56	12:36:05
Scenario 5	8:01:22	9:25:48	10:16:16
Scenario 6	8:23:46	9:47:25	11:12:05

Table 4.11 illustrates the simulation results for part-time incidents. Note that, as the values of non-submitted payments to the system and the value of unsettled payments are all zero, they are excluded from the table. By comparing the simulation result to the benchmark scenario (see Table 4.3) we conclude that, in line with our expectations, more queues and longer delays show up in the system. The average of the total value of queued transactions increased by almost 50% in Scenarios 4 and 6, and by 75% in Scenario 5. The two extremes (minimum and maximum) almost remained unchanged. The average of the maximum queue value also increased notably. Compared to the benchmark case, it became 2.66, 3.55 and 3.85 times higher in Scenarios 5, 6 and 7 respectively. The maximum queue value is more than two times higher in each part-time incident scenario than in the benchmark. The average queue length increased by 12 minutes in Scenario 4, by 27 minutes in Scenario 5, and by 21 minutes in Scenario 6. In the benchmark case the average of the settlement delay equalled 0.07. The average of the delay indicator increased by around 75% in Scenarios 4 and 6, and tripled in Scenario 5.

Table 4.11 **Simulation results for part-time incidents**

	Scenario 4			Scenario 5			Scenario 6		
	Minimum	Average	Maximum	Minimum	Average	Maximum	Minimum	Average	Maximum
Total value of queued transactions (in % of submitted payments)	2.4%	24.5%	37.9%	0.8%	28.7%	40.8%	3.0%	23.5%	37.7%
Maximum queue value (in % of submitted payments)	2.0%	11.4%	24.8%	0.4%	14.4%	31.1%	2.0%	12.2%	24.7%
Average queue length (hh:mm:ss)	0:23:48	0:52:58	1:21:04	0:46:16	1:07:56	1:59:37	0:16:52	1:01:57	1:27:19
Settlement day	0.05	0.12	0.23	0.07	0.2	0.35	0.03	0.12	0.21

If we compare the part-time incident scenarios with each other it is interesting that Scenarios 4 and 6 show very similar properties. In both cases the incidents lasted for four hours; nevertheless in Scenario 4 one, while in Scenario 6 two, banks suffered from the operational incidents. The similar queue and delay properties can be explained by two adverse effects. First, the value of payments submitted later is higher in Scenario 6. This should obviously result in larger queues and longer delays. Second, the banks suffering from the operational incidents also transact intensively with each other. If both of them send their payments as soon as the incidents are sorted out, their payments will not remain long in the queue. This is why it could happen that we do not experience more significant queues and delays in Scenario 6 than in Scenario 4. In Scenario 5 the operational incident lasted for six hours. The simulation results are in line with our expectation – there are longer and larger queues and more significant delays.

4.5 Conclusions

The simulations carried out can be considered a first step in evaluating the ability of the Hungarian payment system to withstand certain types of operational shocks. When entire-day incidents without back-up options and behavioural reactions were simulated, the disturbances to the payment system was severe. In international comparisons, the proportion of unsettled transactions is remarkably high. Queues and delays were also higher than in the simulation exercises of other countries. This can be explained by the high concentration of debit turnover in VIBER and by the fact that the most active participants are generally not those with abundant liquidity. Another important explanation might be linked to the structure and size of the money markets.

If back-up options were employed, then the disturbance to the payment system was significantly less. This stresses the importance and potential efficacy of back-up procedures. However, the shock-absorbing capacity of the system depends heavily on the selection procedure for manually processed payments. In the case of operational incidents, the MNB should stipulate that the banks select not only the transactions with the highest priority but also those with the highest value.

Behavioural reactions were also taken into account; the initially unaffected participants reacted to the operational incident by blocking payments to the stricken bank within two hours. As a result, the value of non-submitted payments increased significantly and the value of rejected payments decreased drastically. By withholding many payments the value of payments remaining unsettled (either not submitted or rejected) was raised. This is the price of isolating the shock and privileging payments sent to other participants. The higher value of unsettled payments means that some of the held-up payments could have been settled.

By analysing the impact of scenarios involving part-time incidents, we concluded that more queues and longer delays showed up in the system. By comparing the part-time incident scenarios with each other, we experienced similar settlement patterns regardless of the number of defaulting banks. The similar queue and delay properties can be explained in this case by the mutual interdependence of the two shocked banks – they trade very actively with each other.

The simulation exercise has several drawbacks, which lead to the overestimation of liquidity risk. First, we assumed that the participants do not raise additional funds. Secondly, we assumed unchanged trading patterns during the entire business day, ie that the banks planned to settle the same volumes and values of transactions with the same counterparties and with the same priorities as under normal business conditions. Probably such behaviour would not take place, as banks are not passive economic agents and would adjust their trading patterns to the shock situation. Banks short of liquidity might postpone some of their transactions and attempt to settle fewer transactions on the day of the shock. Moreover, the banks would surely favour participants with operating infrastructure and with abundant liquidity. Thus, not only would the value and volume of intraday transactions be lower, but also the counterparties to trade with would be reset.

Moreover, not only the trading patterns, but (except for the scenario with the stop sending rule) the settlement behaviour was also assumed to be constant. Banks initially unaffected by the operational incident did not place payments to the technically defaulted bank at the end of the internal or external queues, or did not alter the timing of the payments (for example, by submitting the transaction to the defaulter at the end of the business day). In this way the severity of disruption might be overestimated again. Probably those participants that face liquidity shortage would try to finance payments to the participant under distress from the payments to be received from it by re-prioritizing or altering the time stamps of the transactions.

As we had no information on how the trading patterns and settlement behaviours of banks would change in distressed periods, we tried to keep the simulations as simple as possible and avoided making further, more speculative assumptions. As a next step, by means of qualitative and quantitative research methods, more investigation is needed to model precisely the behaviour of the banks in shock situations.

As for policy implications, the role of the MNB in a crisis situation is of crucial importance. The MNB is the player that communicates with the participants and might be able to influence the settlement behaviour by providing up-to-date information on the nature of the operational incident. In addition, there seems to be a small number of VIBER participants whose operational failure can markedly affect the functioning of the system. In the future, as regards central bank oversight, more attention should be paid to the back-up facilities and procedures of (at least) these critical participants.

References

Arjani, N (2006) **Examining the trade-off between settlement delay and intraday liquidity in Canada's LVTS: A simulation approach.** Bank of Canada Working Paper 2006–20.

Bech, M – Soramäki, K (2005) **Gridlock resolution and bank failures in interbank payment systems.** In: Harry Leinonen (ed.) (2005): Liquidity, risks and speed in payment and settlement systems – A simulation approach. Bank of Finland Studies E:31, 150–177.

Bedford, P – Millard, S – Yang, J (2004) **Assessing operational risk in CHAPS Sterling: A simulation approach.** In: Bank of England's Financial Stability Review, June 2004, 135–143.

BoF (2005) **Description of BoF-PSS2 databases and files version 1.2.0.** Suomen Pankki – Finlands Bank. Financial Markets Department and Research Department (Bank of Finland), Maritta Halonen (MSG Software Oy).

Enge, A – Øverli, F (2006) **Intraday liquidity and the settlement of large-value payments: A simulation-based analysis.** Economic Bulletin 1/2006, Norges Bank.

Financial Stability Review (2007) **ELLIPS: stable ship in stormy waters?** In: Financial Stability Review, National Bank of Belgium, 103–111.

Glaser, M – Haene, P (2007) **Simulation of participant-level operational disruption in Swiss Interbank Clearing.** Presentation held by Martina Glaser at the 5th Payment and Settlement Simulations Seminar and Workshop, 28 August 2007, Bank of Finland, Helsinki.

Kiss, M N – Tanai, E (2004) **Demand for central bank money and provision of credit in the payment systems.** MNB Mimeo.

Lublóy, Á (2006) **Topology of the Hungarian large-value transfer system.** MNB Occasional paper 57.

Lublóy, Á – Tanai, E (2008) **Operational Disruption and the Hungarian Real Time Gross Settlement System (VIBER).** MNB Occasional Papers 75.

Mazars, E – Woelfel, G (2005) **Analysis, by simulation, of the impact of a technical default of a payment system participant.** In: Banque de France Financial Stability Review No. 6, 113–124.

MNB (2001) **Act LVIII of 2001 on the Magyar Nemzeti Bank.** CompLex CD Archive. http://www.complex.hu/. Accessed: 8 December 2003, 14:08.

Schmitz, S – Puhr, C – Moshammer, H – Hausmann, M (2006) **Operational risk and contagion in the Austrian large-value payment system ARTIS.** In: OENB Financial Stability Report No. 11, 96–113.

Tanai, E (2007) **Foreign exchange settlement risk in Hungary.** 2nd report, MNB.

Chapter 5

Simulations in the Dutch interbank payment system: A sensitivity analysis

Ronald Heijmans

5 Simulations in the Dutch interbank payment system: A sensitivity analysis

Abstract

This paper presents an analysis of the sensitivity of the Dutch interbank payment system with respect to value transferred and the amount of available collateral. The Dutch system can be characterised as having a few large and many fairly small participants. Historical data has been used and modified to create a stress scenario. The changes in respect to the historical data are an increase or a decrease in payment values for one of the large participants. These changes in payment value are calculated for the three large banks in the Dutch system. The collateral level was modified between the different stress scenarios. In total, three levels of collateral are investigated, of which one is based on historical data and two on theoretical values that would optimise collateral usage. The results of this paper are presented in terms of both the number of banks affected and the amount of unsettled values at the end of the day. A disruption of one of the large banks does not have a huge impact on the other large banks in the system. The small banks however do face more liquidity problems as a result of the disrupted (large) bank.

5.1 Introduction

Total disruptions of payment systems are relatively rare. Therefore there is not much historical evidence on how payment systems react to a disruption. Even in the current financial crisis the payment systems worldwide have functioned without serious disruptions. One of the most well known operational disruptions in the interbank payment system is the attacks on the world trade centre in New York on 11 September 2001. The massive damage to property and communication systems made it more difficult or even impossible for some banks to execute payments to other banks. Such a disruption can have effects beyond the immediate counterparties of banks disrupted by the shock. In extreme cases they might even disrupt the whole financial system.

Simulations are used to gain a better understanding of the potential impact of a disruption in the interbank payment system. The Bank of Finland has developed a payment system simulator (BoF PSS2) that has been proven to be a very useful tool for gaining insight into the effects of operational disruptions. Many simulations have been run using the BoF PSS2. Ledrut (2007) describes the impact of an operational failure, varying the time at which the disruption takes place and McVanel (2007) investigated the impact of unanticipated defaults in Canada's Large Value Transfer System. A new direction of development of the BoF PSS2 is the integration of network topology; Soramäki et al (2007).

The simulator provides possibilities to simulate a delayed net settlement system (DNS) or a real time gross settlement system (RTGS). A DNS, which settles the payments after a predefined time span by transferring the net positions, uses the smallest amount of liquidity. The settlement risk of a DNS is the highest, because in case of a defaulting participant all payments after the last settled transactions may have to be unwound. A RTGS, which settles payments immediately and individually, requires the largest amount of liquidity, but the settlement risk is very low. There are many system designs between RTGS and DNS, often called hybrids, which require less liquidity than a pure RTGS.

The goal of this paper is to describe the potential impact on the payment system when one large participant in the Dutch interbank system faces a given (small or large) disruption. In the Netherlands, the impact of small-participant disruptions are modest for the total payment system, due to the small payment values involved. Therefore, a disruption of a small participant in the payment system is not discussed in this paper.

The types of disruptions simulated in this paper are an increase and a decrease in the outgoing transaction value of a certain percentage. The potential impact is measured by the number of banks with unsettled payments at the end of the day and by the corresponding value of unsettled payments. An increase in the total transaction value resulting from a temporary increase in the obligations of one large participant means that that participant 'provides' the other participants with extra liquidity (due to increased obligations) up until the large participant runs out of liquidity. The important issue here is how the extra liquidity is distributed over the participants: homogeneously or heterogeneously. A decrease in outgoing transaction value for the large participant leads to a decrease in liquidity available to the other participants. In this case it is interesting to see to what extent the other participants are able to deal with the lower level of liquidity without a disruption of the whole payment system. One important parameter for the amount of available liquidity is the amount of central-bank-approved collateral that is available to the participants. Simulations provide good insight into the potential effects, ie into the payment system's sensitivity to disruptions.

The paper proceeds in a straightforward manner. Section 5.2 describes the stress scenarios that are simulated and lists the general characteristics of the three large banks that are used to create the stress situations. Section 5.3 analyses the number of banks affected given those stress situations, and section 5.4 shows the value of unsettled payments of banks that use all of their collateral for intraday credit, the amount of collateral used by the banks with unsettled payments, and the total negative end of day balance of participants that have settled all their payments in a given stress scenario. Section 5.5 concludes and offers some policy recommendations.

5.2 Data description

5.2.1 The Dutch large value payment system

The Dutch large value payment system, referred to as TOP until 18 February 2008, was a Real Time Gross Settlement (RTGS) payment system. From 1999 to February 2008, it was part of the European TARGET1 system for euro-denominated payments.[1]

For a payment system to function properly it is essential that participants have sufficient funds for making payments without delay. The means of achieving this vary across systems. In the Dutch system DNB,[2] like other Euro-system central banks, provides intraday credit (secured by collateral); see Ledrut (2006) for a discussion of the optimal provision of intraday liquidity. In effect, participants can use as payment liquidity their own credit balances or the credit facility. Participants can thus execute their outgoing transactions before receiving their incoming transactions for that day. Free intraday credit therefore facilitates smooth functioning of the payment system and prevents gridlocks. If the day's closing balance is negative, the participating institution have to pay overnight interest.

---

[1] TARGET: Trans-European Automated Real-time Gross settlement Express Transfer system.

[2] DNB: De Nederlandsche Bank (Dutch Central Bank).

5.2.2 Top data

The data set consists of transaction data for the months December 2005 and April 2006 from the Dutch large value payment system (TOP). It includes all transactions carried out during these two months, both domestic and cross-border. The data set includes both payments made during regular opening hours of TOP and during the evening settlement. The incoming cross border transactions are also included in the dataset. Table 5.1 shows some key characteristics of TOP and compares them to other RTGS systems.

Table 5.1 **Key characteristics of daily payments in TOP, TARGET, CHAPS Sterling and Fedwire**

	TOP[1]	TARGET	CHAPS (Sterling)[2]	Fedwire[3]
Participants	155	10,197	Not available	6,819
Of which direct participants	100	1,126	15	Not available
Transactions (x thousand)	18.4	312	116	519
Value (x billion EUR)	120	1,987	297	1,634
Average transaction value (x million EUR)	6.5	6.4	2.6	3.1

Source: TOP (datawarehouse DNB), TARGET (ECB Bluebook), CHAPS and Fedwire (BIS, 2007 #3053).

[1] The values listed for TOP include only the figures are reported to the ECB for Blue Book statistics. They do not include incoming cross border payments, which do fall within the scope of this paper.

[2] The Pound sterling values are converted via US dollars to Euros.

[3] The US dollar values are converted to Euros via the exchange rate EUR/USD 0.8051, which is the average exchange rate for 2005 as listed in Red Book (BIS) 2006.

Some participants have more than one account in TOP. These accounts are used by the participant for administrative purposes. In the simulations these accounts are treated separately. The results are however for the participant level. In case several accounts of a single participant have unsettled values at the end of the day, as a result of the disruption, they are treated as one account.

5.2.3 Stress scenarios

A stress scenario in this paper is defined as a modified dataset, based on historical data, given certain conditions. There are two parameters in the historical payment datasets, which have been modified in defining the stress scenarios:

– the outgoing payment value of one participant, expressed as a percentage (8 alternative values: 50%, 75%, 90%, 95%, 110%, 125%, 150% and 200%),
– available collateral values for all participants; 3 alternative values: historic available collateral for intraday credit (HC), minimum liquidity to settle all payments at the end of the day when there is no disruption (lower bound or LB), and the liquidity necessary for immediate settlement of all payments when there is not disruption (upper bound or UB).

The modified payment values are applied to:

– three large banks one at a time,
– two different months.

The disruption period is set to:

– two different lengths of time (one day or the whole month).

To define a scenario one option from each of the five items is chosen.

The three chosen banks are the three largest banks in TOP with respect to value transferred per day. For each of these three banks, the outgoing payment value either increased or decreased by a certain percentage. A decrease in payment value could result from a technical problem or an intentional delay by the participant. An increase in payment value could result from paying off of loans or a temporary increase in obligations. It might also be that customers move funds to other banks because they no longer trust their bank, due to market rumours about the bank, as observed in the current crisis. This may also be seen as a decrease in the payment value of all other banks. Even though this increase in outgoing value provides extra liquidity to the other participants as a whole, it need not mean that each participant will get its own share. The interesting question concerning such an increase in liquidity is whether this liquidity is distributed evenly over all participants.

The data of December 2005 and April 2006 were used to define stress situations. The reason for choosing two different months is to investigate differences in outcome of stress simulations. It would be desirable from a statistical point of view to investigate all the months of the year, but this is a rather time consuming operation. The length of the disruption is either one day or the whole month. A disruption of one day means that unsettled payments at the end of the day will be deleted before going to the next. When the disruption or behaviour of a single bank lasts the whole month unsettled payments at the end of the day are the first payments to be settled the next day. Even though a disruption of a month is not very likely it provides a measure of how long a disruption needs to last before crucial banks in the payment system are also affected. In other words, is two days sufficient for a total disruption of the payment system or does this perhaps require 20 days or more. The more time it takes for a total disruption, the more time central banks have to take counter-measures.

5.2.4 Payments statistics of the three large banks

Figure 5.1 shows the total payment flows in per cent of total outgoing payments in December 2005 for the three large banks. The total value of outgoing payments excludes incoming cross border payments. An interesting aspect is the net payment flow between participants. Given the fact that in the long run there cannot be net payers or receivers in the system, this extra flow between two participants must come back from other participants (domestic or cross border); see eg Pröpper et al (2007) on network characteristics. The figure also shows that the three large banks have a lot of cross border payments. This is important because this paper looks at the potential impact of a disruption on the domestic market. The more a bank pays cross border, the smaller the potential impact on the domestic market.

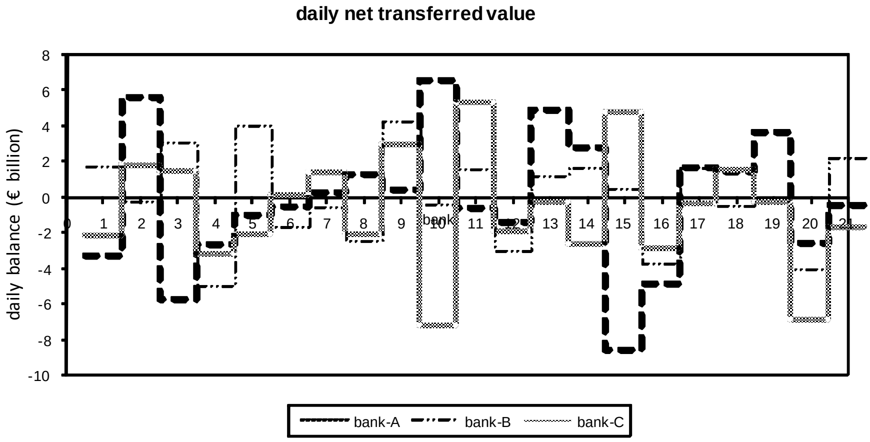
Figure 5.2 shows that there is wide variation in the daily net value transferred for the three large banks in December 2005. The daily balances of bank-A vary between EUR -9 and EUR +7 billion. The balances of bank-B and bank-C are slightly lower than that of bank-A.

Figure 5.1 **Payments flows between the three large banks and between the large banks and other domestic participants and cross border payments in December 2005.**[3]

Payment values are in % of total outgoing payments of TOP participants (this excludes incoming cross border payment values), which is equal to EUR 2556 billion.

[3] The payment transactions from and to the same participant are not shown.

Figure 5.2 **Daily net transferred values for the three large banks in December 2005**



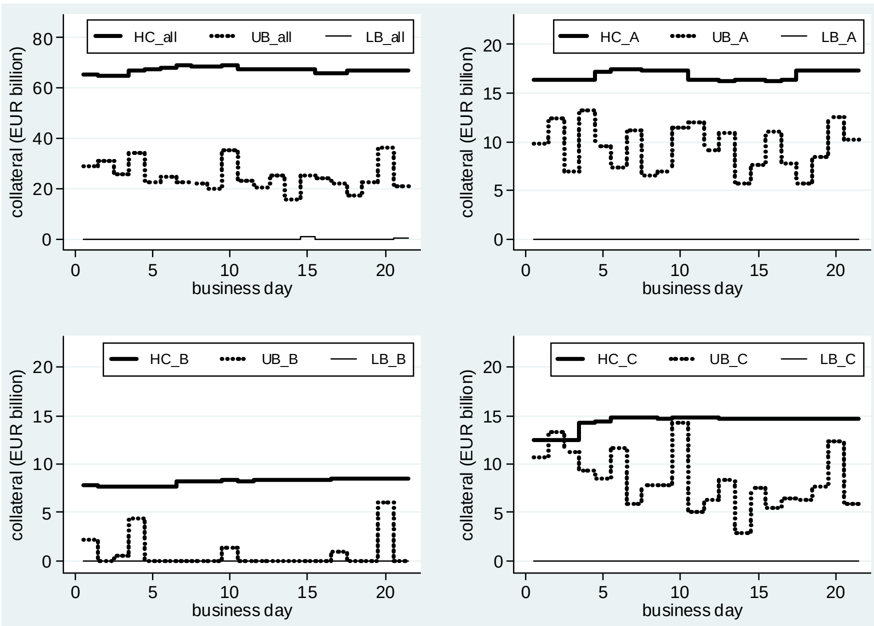
5.2.5 Collateral values

Simulations were run with three different values of collateral; see Section 5.3.1. The HC collateral level shows the maximum level of intraday credit a participant can obtain. The UB and LB are theoretically calculated levels of collateral. It would be difficult to determine these levels prior to the start of a new business day because it is not (exactly) known at what time and how many payments will be paid and received that day. If a bank faces an operational disruption and cannot pay the planned amount, it is likely that some of the other banks will run out of liquidity (including intraday credit) and cannot execute all payments immediately in the case of the UB or by the end of the day in the case of the LB. If the disruption is large enough, there will even be unsettled payments at the end of the day, when the UB collateral level is used.

Figure 5.3 shows the three different collateral values of bank-A, bank-B, bank-C and all participants together for December 2005. The HC (thick solid line) shows more fluctuation in the values, which means that these banks participated in open market operations during the month. The upper bound (dotted line) fluctuations vary for the three large banks. Bank-A shows values between EUR 6 billion and EUR 14 billion. Bank-C shows values between EUR 0 and EUR 6 billion. A value of EUR 0 for the upper bound means that the banks did not need any collateral on these days to settle all their payments immediately. The values for bank-C fluctuate the most: between EUR

2 billion and EUR 14 billion. On a few days of this month the value of the upper bound is higher than or close to the maximum collateral value. This suggests that bank-C actively uses its collateral. The lower bound is zero for the three banks and almost zero for all participants. This is due to the fact that all banks have an obligation to hold a certain level of credit on their account for their minimum reserve requirement. These funds can be used to settle payments intraday while at the end of the day their balances are levelled to meet the requirements.

Figure 5.3 **Collateral values for all banks together (top left), bank A (top right), bank B (bottom left) and bank C (bottom right)**



The total value of the HC varies between EUR 44 billion and EUR 59 billion. For HCex, the variation is larger than for the HC. The variation of the UB, between EUR 20 billion and EUR 36 billion, is comparable in absolute value to the HCex but in relative terms it varies more. The LB, on the other hand, shows values close to zero most of the time. This means that most banks will be able to settle their obligations by the end of the day without having any collateral.

From Figure 5.3 it can be seen that there is more collateral available in the payment system as a whole than is needed. This means, in theory, that there is enough collateral available to execute all payments without any delay. However the distribution of the collateral is not optimised over the participants. For the three large banks there are several days in which the UB is higher than the HC.

5.2.6 Simulations and assumptions

The scenarios described in section 5.3.2 are based on several assumptions, which do not necessarily reflect what happens in reality. They are necessary for studying the potential impact of a disruption given a certain stress scenario. The assumptions are:

– no extra liquidity from other accounts
– no extra collateral can be pledged
– participants cannot borrow from other banks
– participants do not react in any way to the stress situation.

5.3 Results: number of banks affected

5.3.1 Introduction

This section gives the results of the simulations with respect to the number of banks affected by the disruption. The figures presented in this section refer to December 2005. The results for April 2006 are very similar. In this paper, to be affected means that a bank has unsettled transactions at the end of the day. Also banks which do not have unsettled payments are impacted by the disruption but are able to 'absorb' the shock and to fulfil their own obligations. The number of banks affected gives insight into how the disruption spreads across the payment system but says nothing about the value of unsettled payments, which will be described in section 4. The description of the different scenarios was given in section 2.

5.3.2 Bank-A

Figure 5.4 shows the average number of banks affected and the standard deviation, for the 50% to 200% cases using the HC. The number of banks affected, running from 75% to 95%, does not show a decrease of the average number. This is the result of small banks not having any collateral. In case they do not receive some payments, this will automatically lead to unsettled payments at the end of the day. If there is a bank affected in the 110% to 200% cases, it is at least bank-A. This is because bank-A will run out of its liquidity (including its intraday credit) first. In the 200% case, there are other banks affected besides bank-A. This is the result of the inhomogeneous distribution of liquidity. A large part of the liquidity will leak across the border.

Figure 5.5 shows the average number of banks affected for the 50% to 95% and 110% to 200% cases using the UB; see Figure 5.3. The difference between UB and HC varies between EUR -3 and EUR 8.5 billion. The average number of banks affected for the 50% and 75% cases is higher for UB than for HC. This is the result of the lower collateral levels for most banks in the UB than in the HC. Due to the lower collateral value of bank-A, it will run out of liquidity more quickly for the 110% to 200% cases. This means that bank-A is not able to absorb the same shock in the same way as in the HC.

Figure 5.6 shows the average number of banks affected for the 50% to 95% cases and 110% to 200% cases using the LB. The average number of banks affected in both figures is clearly higher than in the previous two collateral cases. For the 50% case, the average number of banks affected rose from 7 to 16 banks for the 200% case. For the HC scenario, the average number of banks rose to 9 for the 50% case and 3 for the 200% case. From this figure one can conclude that under 'normal' circumstances the Dutch payment system does not need collateral to execute all payments by the end of the day, but a small disruption by bank-A results in many banks being affected by the end of the day. In other words, the shock absorbing ability of the payment system declines to a low level. For the multiple day scenario, the number of banks affected rises to 52 banks for the 50% case and to 33 for the 200% case.

The multiple day scenarios show an increasing trend of banks affected on almost all days for all cases (95% to 200%). The maximum number of banks affected for the 50% case rises to 34 and for the 95% case to 12 for the HC, ie 46 and 22 banks for the UB and 54 and 30 for the LB scenario, respectively. There will be up to 7 banks for the 110% case and 11 for the 200% case, ie 11 and 26 banks for the UB scenario and 24 and 34 for the LB.

Figure 5.4 **Average number of bank affected when bank-A pays 50%–95% and 110%–200% of its payments, with each day is treated individually and historical collateral amounts are used by all banks**

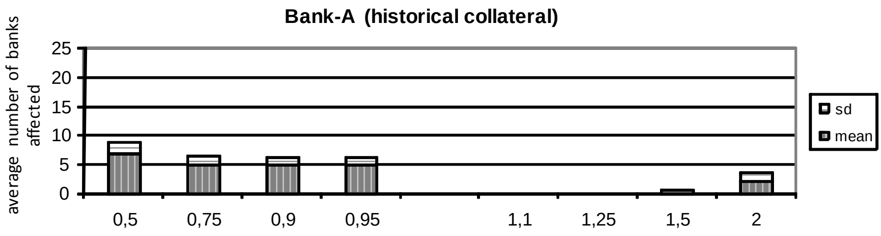


Figure 5.5 **Average number of bank affected when bank-A pays 50%–95% and 110%–200% of its payments, with each day treated individually and the upper bound collateral amount being used by all banks**

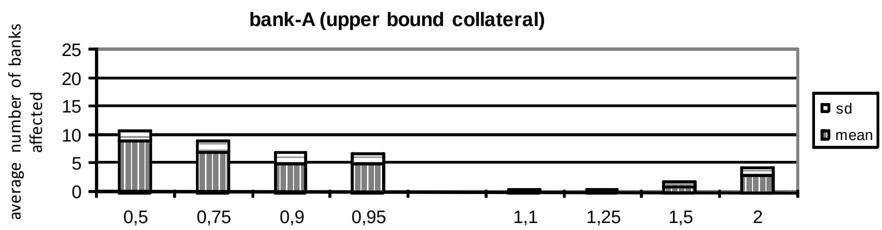
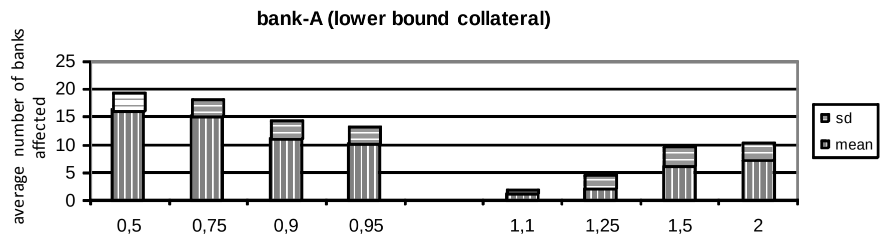


Figure 5.6 **Average number of bank affected when bank-A pays 50%–95% and 110%–200% of its payments, with each day treated individually and the lower bound collateral amount being used by all banks**



5.3.3 Bank-B and bank-C

Figure 5.7 shows the average number of banks affected for the 50% to 95% and 110% to 200% cases for the HC of bank-B. For the 50% case, there are on average 6 banks affected by the shock. The average number of banks affected is just below that for the same disruption of bank-A. This is due to the smaller daily payment values by bank-B. When bank-B increases its outgoing values (110% to 200% cases), there are more banks affected on average compared to bank-A. Even though bank-A has a larger daily payment value than bank-B, the HC is relatively lower for bank-B than bank-A. As a result, bank-B will run out of its liquidity sooner and therefore will impact other participants as well.

The LB collateral of bank-B is zero for all days and the upper bound collateral is zero for most of the business days, which makes these two scenarios very similar. However, two of the days have collateral values of EUR 4.5 and EUR 6 billion. The average number of banks affected for the single day scenario is just over 12 banks for the 50% case, see Figure 5.8, which is approximately twice the number for the HC collateral scenario. The multiple day scenario rises to 43 banks for the 50% case and 24 banks for the 95% case. For the 200% case, the number of banks affected rises to 35. The reason for this is that liquidity leaves the domestic market and flows to cross border TARGET participants. It is assumed in this simulation that cross border participants cannot run out of liquidity.

Figure 5.9 shows the average number of banks affected for the 50% to 95% and 110% to 200% cases for the LB of bank-C. This is lower than for bank-A and bank-B. Even though bank-C has a larger total outgoing payment value (20% of total) than bank-B (13%), it has fewer banks affected for the HC. This is mainly due to the fact most of the extra payment value of bank-C goes to cross border participants. The average number of banks for the UB is only slightly higher because the average amount of collateral in the UB is slightly lower than for the HC. The average number of banks affected for the LB is 3 to 4 times as large as for the HC, but lower than for bank-A and slightly lower than for bank-B.

The multiple day scenarios show the same trend as for bank-A. The maximum number of banks affected is however lower for both bank-B and bank-C. The maximum number of banks affected for the 50% case is 25 for bank-B and 20 for bank-C. For the 200% case this is 11 and 9 banks respectively.

Figure 5.7 **Average number of bank affected when bank-B pays 50%–95% and 110%–200% of its payments, with each day treated individually and historical collateral amounts are used by all banks**

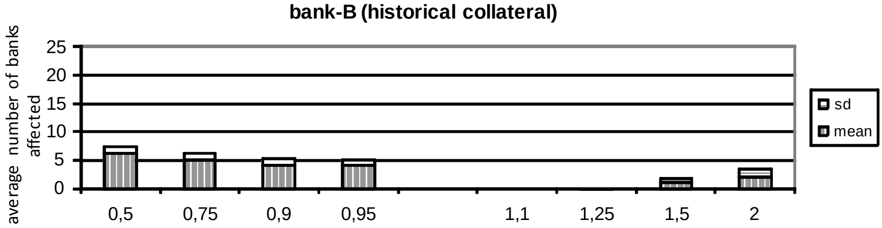


Figure 5.8 **Average number of bank affected when bank-B pays 50%–95% and 110%–200% of its payments, with each day treated individually and the lower bound collateral amount used by all banks**

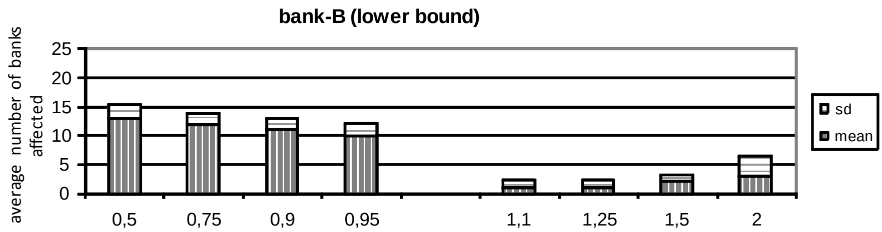
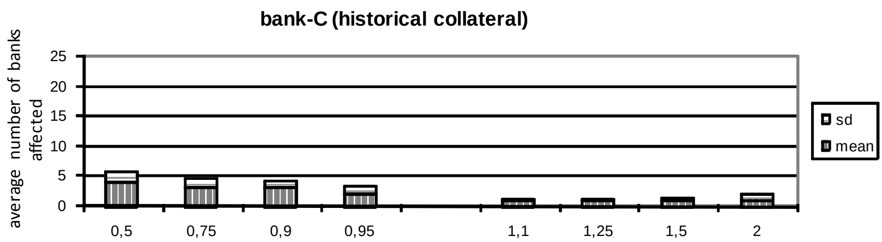


Figure 5.9 **Average number of bank affected when bank-C pays 50%–95% and 110%–200% of its payments, with each day treated individually and historical collateral amounts used by all banks**



5.4 Results: unsettled values

5.4.1 Introduction

The previous section showed the number of banks affected for the different scenarios. The number of banks affected indicates how a disruption spreads across the whole payment system. When only a few (small) banks are affected by the disruption, the impact on the payment system is very limited. However, when many banks are affected, especially large banks, the impact of the disruption on the whole payment system is greater.

The number of banks affected does not say anything about the unsettled value and negative end of day balances which correspond with this number. In a hypothetical case, a bank is EUR 0.01 or EUR 100 billion short for making a payment and this bank will be 'affected' according to the definition in section 5.3. The first case however will be much easier to resolve than the latter one. This shortage of liquidity can be resolved by eg an incoming payment or bringing in fresh collateral. The effect of bringing in extra collateral is not investigated in this paper; see the assumptions in section 5.2.6.

This section describes three different values as a result of the disruption:

– the mean value of unsettled payments at the end of the day
– the mean collateral of banks with unsettled payments at the end of the day
– the mean negative end of day balances of banks without unsettled payments.

The mean value of unsettled payments indicates the direct impact of the disruption on the payment system. The mean collateral of banks with unsettled payments provides information on the type of bank. The banks with large values of collateral are relatively large ones while banks with little collateral are usually relatively small. There are also participants which do not have any collateral at all. The mean negative end of day balance of banks without unsettled payments indicates the impact of the disruption on the other participants.

5.4.2 Bank-A

Figure 5.10 and Figure 5.11 show the values of the 50% to 95% and 110% to 200% cases of the HC. From the figures, it can be seen that the amount of collateral of banks with unsettled payments at the end of the day (red bars) is almost zero. These banks are usually very small and in most cases subsidiary banks of larger banks which do have significant amounts of collateral. These subsidiary banks are in most cases funded by the parent bank in the morning, and at the end of the day the liquidity is transferred back to the parent. It can also be seen from the figures that when the disruption of bank-A is small, say 95%, the other banks are able to absorb the shock without any problem. This does not mean that banks which are able to absorb the shock are also willing to absorb it. When a bank suspects that the disrupted bank, in this case bank-A, has serious financial problems, it may stop sending payments even though it has sufficient liquidity to make a payment. The results for the UB are similar. This is the result of the low (almost zero) collateral values of the affected banks.

Figure 5.10 **The average values of bank-A using historical collateral amounts for 50%, 75%, 90% and 95%, with each day treated individually.**

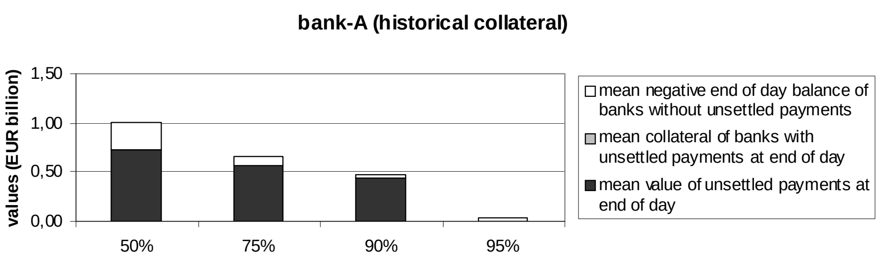
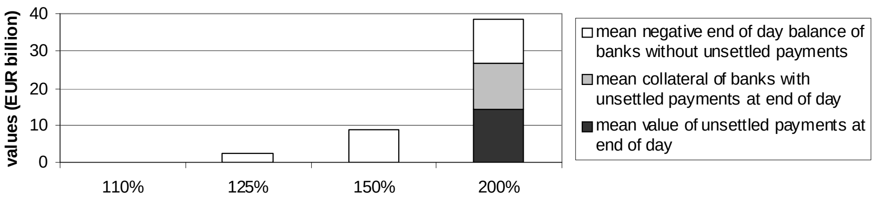


Figure 5.11 **The average values of bank-A using historical collateral amounts for 50%, 75%, 90% and 95%, with each day treated individually**

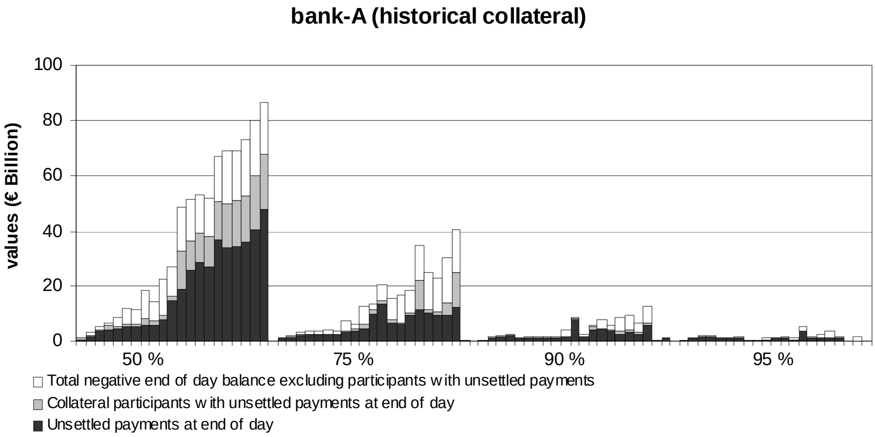


The values of the LB are twice as large as the HC for the 50% to 95% cases and roughly the same for the 110% to 200% cases. The values of the LB consist only of unsettled payments at the end of the day (blue bar). This is because there is hardly any collateral available and therefore banks cannot obtain any intraday credit. In the Euro system, banks must secure intraday credit against eligible collateral.

Figure 5.12 shows the values of the 50% to 95% cases for the multiple day HC scenario. From the 50% case, it can be seen that the sum of the values shows an increasing trend if the shock continues for the whole month. From the 14th day, the sum of the three values jumps. This is mainly due to the used collateral, which increases by a factor 10 to EUR 16.2 billion. A large part of this collateral is from bank-C, which for the first time this month has unsettled values at its main account. The increasing trend is also visible for the 75% and 90% cases but not as strong as the 50% case. For the 95% case, this trend cannot be observed.

The 110% to 200% cases show increasing trends for all four cases. For the first few days of the 110% and 125% cases, all values are settled by the end of the day. This can also be seen in Figure 5.4, which shows the number of banks affected by the disruption. The trend for the number of banks affected is not the same as for the values. There is both an increase and a decrease in the number of banks affected at the end of the month. This means that even though the maximum potential impact increases in value, the number of banks affected does not\. Most of the unsettled values are to bank-A. However, an increasing amount is linked to other participants. This can be seen by the increasing use of collateral, from EUR 15.9 to EUR 30.9 billion for the 200% case.

Figure 5.12 **The values of bank-A using historical collateral amounts for 50%, 75%, 90% and 95%, when the disruption continues for the whole month**



5.4.3 Bank-B and bank-C

The average values of the 50% to 95% cases for the HC of bank-B are similar to those for bank-A. However, the differences in values between the different days are much smaller than for bank-A. The average values for the 125% to 200% are lower than for bank-A. The average daily outgoing payment value of bank-B is half the value of bank-A. The impact of the disruption of bank-B is however similar. This is the result of small banks being affected. For these banks, even a small decrease in the amount of incoming payments causes payment problems. As for the disruption of bank-A, most banks which are affected are relatively small and do not have much or any collateral and are often subsidiaries of larger banks. The LB and UB of bank-B show values which are approximately 50% of the value of bank-A.

When the shock continues for more than one day, the total impact (the sum of the three values) increases over the days. The 50% to 95% cases of the HC MD scenario show an increasing trend. This trend is not as clear as for bank-A situation. The total value for the 50% case is significantly lower than in the bank-A case. This observation is not surprising in light of the total amount of outgoing payments in December 2005, listed in Figure 5.1. Secondly, the amount of cross border payments is relatively higher for bank-A (17.1%) than for bank-B (6.5%) and bank-C (12.5%). For 110% to 200%, the trends are very similar to those of bank-A.

The total payment value of bank-C is smaller than bank-A, but larger than bank-B; see Figure 5.1. The average value of the payment of bank-C is however much higher than the other two, EUR 21 billion compared to EUR 9 billion for bank-A and bank-B. This is an important characteristic in case of a disruption. In situations in which there is not much liquidity available on the accounts a large value becomes more difficult to settle than smaller ones. This is especially important in the 200% case because the payments are doubled and therefore even larger.

The payment values of bank-C and bank-B for the domestic part of the payment system are almost equal while the value for bank-A is significantly larger. This explains why the impact of bank-B and bank-C show similar results for the 50% to 95% cases. For the decreased value cases, only the domestic payments are relevant because the cross border TARGET participant (by assumption) will never face payment problems.

5.5 Conclusions and policy recommendations

This paper has demonstrated the sensitivity of the payment system to a disruption of a large bank.

1. The Dutch large value payment system shows a high level of resistance.

2. Intraday credit is an important aspect of a large value payment system. The more collateral banks have, the more intraday credit they can obtain. When banks do not have any collateral, they are still able to fulfil all their obligations at the end day under normal circumstances, but the intraday queues will be longer. In times of (small or large) disruptions, there will be banks which cannot fulfil their obligations at the end of the day. This argues for substantial amounts of collateral such that temporary technical disruptions will not automatically lead to long queues and unsettled payments at the end of the day. The ability to bring in extra collateral to obtain more intraday credit can help to reduce the impact of disruptions. A smooth functioning collateral management system is therefore required.

3. Between January 2006 and 2009, the total amount of collateral increased by nearly almost 200%! A part of this extra collateral was used for open monetary loans, but a large share collateral was still available for extra intraday credit. The banks have not used significantly more intraday credit even though they were able to do so. This means that all banks have become better able to absorb shocks.

4. It is important to realise that this paper discusses the ability to withstand a disruption. This does not mean that banks are willing to continue making payments in cases of disruption. This depends on the cause of the disruption. When banks fear that a certain bank is unable to meet its obligations as a result of financial problems, as in the current crisis, they may delay or even stop making payments for some time to the disrupted bank or, in an extreme case, to all other banks.

5. The affected banks are mainly small banks which do not have any or only a small amount of collateral. This means that if they do not receive some payments, this will automatically lead to unsettled payments at the end of the day. These small banks are quite often subsidiaries of larger banks. In case they do not have sufficient liquidity to make their payments, the parent bank can supply them with extra liquidity. This will reduce the impact of the disruption.

6. For the domestic payment system, the disrupted bank’s level of cross border payments is important. When a bank cannot execute a part of its payments, the impact will be less if the bank has more cross border payments and vice versa. The increase of cross border payments within the European Union over the last decade has reduced the influence of local authorities such as supervisors and overseers.

References

Bank of international Settlement (2006) **Red Book: Statistics on payment and settlement systems in selected countries, figures for 2005.**

European Central Bank (2006) **Blue Book: Payments and Securities Settlement systems in the European Union and in the acceding countries, addendum incorporating 2005 data.**

Ledrut, E (2006) **A tale of the water-supplying plumber: intraday liquidity provision in payment systems.** DNB working paper series 99.

Ledrut, E (2007) **Simulating retaliation in payment systems: Can banks control their exposure to a failing participant?** DNB working paper series 133.

McVanel, D (2005 and 2006) **The impact of unanticipated defaults in Canada's Large Value Transfer System.** Proceedings from the Bank of Finland Payment and Settlement Systems Seminars.

Oord, A v – Lin, H (2005) **Modelling inter- and intraday payment flow.** DNB working paper series 74.

Soramäki, K – Beyeler, W – Bech, M – Glass, R (2005 and 2006) **New approaches for payment system simulation research.** Proceedings from the Bank of Finland Payment and Settlement Systems Seminars.

Chapter 6

Structure and stability in payment networks – A panel data analysis of ARTIS simulations

Stefan Schmitz – Claus Puhr

6 Structure and stability in payment networks – A panel data analysis of ARTIS simulations

Abstract

The purpose of this study is to investigate the impact and importance of network structure, at the network and node levels across days and scenarios (stricken ARTIS participants) for the stability of payment systems hit by operational shocks. The analysis is based on a large number of simulations of the Austrian large-value payment system ARTIS that quantify the contagion impact of operational shocks at participants' sites. We find that only a few payment system participants are systemically important and that contagion displays substantial variation across time and across scenarios. A subsequent panel data investigation is aimed at explaining the variation across time and network participants by structural differences in the payment network across time and the position of the stricken account within the network. It shows that (i) standard variables such as liquidity and liquidity loss can explain a substantial fraction of the variation across both time and scenarios, (ii) the structure of the network itself adds very little and (iii) the position of the stricken account within the network indeed contributes moderately to explaining the variations in contagion. Relative explanatory power is higher when the analysis focuses on contagion measured by the number of banks with unsettled payments or the value of unsettled payments than when the measure is based on the number of unsettled payments. Because the structural indicators add very little explanatory power to the more traditional measures of the role of an individual participant in the payment system (value and volume of payments), we conclude that at this stage network indicators seem to be of limited use for stability analysis.

6.1 Introduction

Recent work on the stability of banking systems suggests a systematic relationship between network structure, system stability and contagion.[1] Similarly, a recent study conjectures that network structure might be relevant for the stability of payment systems.[2] In previous research we found wide variation in contagion impacts of an individual bank's failure to process payments across banks, across days, and across scenarios.[3] Here we investigate whether the position of the stricken bank within the network helps to explain contagion across scenarios and whether daily variations in network structure contribute to understanding the contagion variation across days.

Studies concerning network stability in the real world[4] have focused on the often observed fact that a few nodes have a large number of links, while most nodes have only a few. The reason that so much attention has been placed on these 'scale-free networks'[5] is their robustness to random node removal (the common way of assessing instability). However, a targeted attack in which the most highly connected nodes are removed leads to rapid disintegration. In financial stability analysis, this framework and focus might be relevant for interbank credit (where establishing a credit relation – a link – is costly). The physical network structure of ARTIS, however, is not scale-free but complete.[6] Thus connectivity is not the relevant stability concept.

The stability problem in ARTIS is not that bank A cannot make a payment to bank B because the banks are no longer linked. The problem is that bank A might not have enough liquidity to make the payment because it has not yet received payments from other banks in the system. As connectivity relates to the flow of liquidity in the system and the liquidity flows through central nodes exceed those through the peripheral nodes, it plays an indirect role for the analysis of stability. Therefore, our measures of the contagion impact of shocks

---

[1] Inter alia Boss et al (2004).
[2] Soramäki et al (2007).
[3] Schmitz and Puhr (2007).
[4] Eg the Internet, but also large value payment systems such as FedWire and BOJ-NET, or the Austrian interbank market.
[5] Scale-free networks are a special case of the aforementioned networks with few important and many minor nodes (in terms of links), where their degree distribution follows a power law $P(k) \sim k^{-\gamma}$.
[6] Participants need not submit payments to each other via hubs; they can do so directly. The only exception would be a failure of the entire payment system infrastructure, but this question is beyond the scope of this paper.

focus on the impact of the shock on the flow of liquidity (ie unsettled payments) rather than on the disintegration of the network.

To quantify the contagion effect following the failure of an individual bank, we run about 30 000 simulations based on actual ARTIS transaction data for the period 16 November 2005 to 16 November 2007, following the methodology presented in our previous work on ARTIS.[7] In addition to this quantification of (contagiously) unsettled payments in case of an operational incident, we calculate a large number of network indicators on the network (44) and node (71) levels for each scenario (stricken ARTIS participants) and for each day in the sample.

In the main body of this paper we investigate whether the variation in network indicators can explain the variation in contagion. We start with a univariate analysis at the network-level on variation across days and at the node level on variation across days and scenarios (stricken ARTIS participants). In a second multivariate step, we conduct an exploratory panel data analysis which includes network indicators at both levels and we show how well they explain the variation in our three measures of contagion across scenarios and days (number of banks with unsettled payments, volume of unsettled payments, and value of unsettled payments).

The remainder of the paper is structured as follows. In section 6.2 we present data on the network structure of ARTIS. Section 6.3 introduces the simulations. Based on the results, we discuss the scale of contagion in ARTIS and try to provide a means to determine systemically important banks. Section 6.4 covers the univariate analysis and provides a first glance at the relation of network and node-level indicators to the contagion effects in the simulations. In section 6.5 we cover the multivariate analysis and present the results of our panel data analysis. Section 6.6 wraps up our findings.

6.2 Measures of network structure[8]

The definition of the network under investigation is not trivial in empirical network analysis. We focus on the Giant Strongly

[7] Schmitz and Puhr (2007) and Schmitz et al (2008).

[8] For detailed definitions, formulas, and graphical illustrations of the network indicators, see Appendix 2.

Connected Component (GSCC) of ARTIS.[9] The GSCC is the largest component of the network, in which all nodes connect to each other via directed paths.[10] We have chosen this definition of the network for two reasons: first, ARTIS contains a comparatively large number of accounts which are not related to financial stability (ie offset accounts of OeNB's cash distribution subsidiary) and which are not active on most of the days in the sample. Second, we want to ensure the comparability of our data with that reported for FedWire in Soramäki et al (2006) which refers to the GSCC.[11]

A related question is about the selection of the appropriate indicator of network structure, as the number of available indicators is large. At the network-level we calculate 44 network indicators.[12] Similarly, the number of available indicators at the node-level comes to 71. We composed our set of indicators to include those used in comparable studies as well as those suggested by the underlying theory for selecting appropriate indicators for specific typologies of payment flows.

Boss et al (2004) relate contagion in the interbank market to betweenness centrality at the node-level, because this measure has a higher explanatory value than the alternative network indicators in their data set. They find a dented linear relationship. Banks with betweenness centrality $0 \leq C_B(h) \leq 2$ do not cause any contagious defaults. For $C_B(h)>2$ they find a linear relationship with a slope of about 0.8.

Borgatti (2005) studies the selection of the appropriate centrality measure for various typologies of flow processes. He classifies flows along two dimensions: the characteristics of the route through the network and the characteristics of the transfer mode. The first dimension considers the constraint on the sequences in which links and nodes are (repeatedly) passed. Liquidity can be transferred to any other node in the network (including the submitter of the first

[9] For comparable data on the network of all active accounts see Schmitz and Puhr (2007). For a description of the Austrian banking system see OeNB and FMA (2004) The Austrian Financial Markets, Vienna, pp. 50–55.

[10] A directed path is a path that connects nodes without passing any node or link more than once.

[11] For a comparison as well as a more detailed account of ARTIS network indicators, see Schmitz et al (2008).

[12] This includes the directed and/or value/volume weighted and/or average/maximum values of select indicators. Kyriakopulos et al (forthcoming) find a strong dependence of network characteristics on aggregation time. The large number of network indicators and critical role of aggregation time raise the problem of data-mining in network topology studies.

payment). Hence, it is unconstrained (referred to as a walk). The second dimension refers to the way in which the flowing good is passed along the route from one node to another. In the case of liquidity, the initial holder has to part with it (referred to as a transfer).

What does this imply for the flow of liquidity in ARTIS? In a physically complete network banks need not make payments to other banks via third parties. They transfer directly to the ultimate receiver. However, the flow of liquidity does not stop there. Where it ultimately ends up, is beyond the control (and interest) of the initial submitter of a payment. Given that betweenness centrality is based on the share of all shortest paths through a node, it is not a good measure of centrality in a study of liquidity flows. Degree centrality is more suitable for this purpose.

Besides considering the most meaningful indicators we want to ensure a high degree of comparability with other papers and therefore employ a wider range of network indicators. Moreover, we want to investigate whether network indicators in general add value to the more traditional measure used in comparable simulation studies (ie size of the individual node in terms of value and volume of transactions). Therefore we focus on the measures value and volume as well as on the network indicators average path length, degree, connectivity, clustering, betweenness centrality and dissimilarity index as provided in Table 6.1 for the network-level averages across participants.

Table 6.1 **ARTIS network indicators (Network-level)**

	Mean	Median	Min.	Max.	Std.Dev.
Payments					
Volume	15380	15436	9786	25000	2019
Value (EUR bn)	48.5	46.9	22.6	84.9	10.6
Average (EUR mn)	3.20	3.00	1.90	5.90	0.70
Size					
Nodes	133.2	132	112	159	9.3
Links	1376	1376	1222	1602	69
Distance measure					
Avg. Path Length	2.4	2.4	2.2	2.6	0.08
Connectivity					
Average degree	15.6	15.5	14.2	17.8	0.6
Connectivity (%)	7.9	7.9	5.9	9.9	0.8
Clustering (%)	58.3	58.3	51	63.7	2.3
Others					
Betweenness cent. (%)	0.8	0.8	0.6	0.9	0.1
Dissimilarity index	0.47	0.47	0.39	0.60	0.03

Source: Own calculations based on daily averages of the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays).

For our observation period, 16 November 2005 to 16 November 2007, the average volume of transactions per day is 15380 in the GSCC of ARTIS. The average value of transactions per day is EUR 48.5 billion and the average transaction size is EUR 3.2 million. The size of a network is defined by the number of nodes. On average, there are 133.2 nodes in the GSCC during the sample period, of which 63 are in the GSCC on all days. The active nodes are linked by an average of 1376.1 directed links.[13]

An indicator of distance between nodes is the shortest path length. We calculate the average shortest path length for each originating node by averaging over terminating nodes and then averaging over originating nodes, to get the average path length of the entire network. Across days, it is 2.4.

How well nodes are connected in the network is captured by the average degree of the network. It is calculated by summing over all (undirected) links originating from the nodes and then averaging over nodes.[14] Averaged also across days, it amounts to 15.6 in the ARTIS

[13] The average number of nodes in ARTIS active each day was 209.8 and the number of directed links was 1637.5.

[14] The out-degree refers to the number of links originating at the node while the in-degree is based on to the number of links terminating at the node. Over the network, the average out- and in-degree are equal to m/n.

system. However, the most active nodes have many more links originating and terminating at them.[15] The connectivity of a network is captured by the number of actual links relative to the number of possible links. Connectivity averages 7.9 per cent. The clustering coefficient provides a measure of the average connectivity of the neighbours of all nodes in the GSCC. On average, about 58 per cent of neighbours of the nodes are also linked.

Betweenness centrality measures how many shortest paths pass through the average node. The 8 per cent figure is quite low and stems from the centrality of a few nodes with high betweenness centrality and a large number with low values. Finally, the dissimilarity index captures the relative viewpoints of the network from two neighbouring nodes. If the network looks very similar from both nodes, the dissimilarity index number is low. In the GSCC it is 0.47, which implies that on average the perspectives of the GSCC differ substantially from any two neighbouring nodes. A lot of nodes are linked that otherwise do not share many network characteristics. In sum, we interpret the data on network indicators as corroborating previous evidence that a few large nodes dominate the payment system and that many of the smaller nodes are connected to the largest nodes at the centre of the network.

6.3 Measures of network stability

As argued in the introduction, connectivity is not an adequate criterion to capture the systemic impact of an operational problem at one of the nodes in a large value payment system. Alternatively, we suggest defining a threshold based on the average contagion effect following the failure of an individual payment system participant. As operational failures, let alone such with systemic impact, are few and far between, we resort to simulations. These provide us with what we call contagious defaults, which can be measured in three ways.

First, the number of participants (banks or transfer accounts) with unsettled payments at the end of day measures how many participants (banks or transfer accounts) faced liquidity problems due to an

[15] For the analysis of the degree distribution, see Schmitz et al (2008), where the hypothesis of a Power Law distribution is rejected for the monthly network and Kyriapopulos et al (forthcoming) who find that degree distributions seem to have a Power Law distribution in daily networks (in ARTIS), but that this property vanishes in longer aggregation times.

operational incident at another participant. Second, the number of unsettled payments at the end of day is the total volume of all payments that could not be settled by participants that were not subject to an operational incident. Third, the value of unsettled payments at the end of day is the total value of all payments that could not be settled by participants that did not experience an operational problem.

We conduct 31311 simulations based on 63 different scenarios for 497 transaction days from 16 November 2005 to 16 November 2007 (excluding Austrian holidays) which yield some 620 million simulated transactions.[16] These simulations are calculated with a self-implemented Matlab-based software tool (inspired by the Bank of Finland Payment System Simulator), tailored for ARTIS.

For this paper we run simulations for all 50 banks in the GSCC on all Austrian working days throughout the sample period and all 13 Transfer accounts[17] that form part of the system on all days in the sample period. We assume an operational incident that hits one participant (banks or transfer accounts) in each simulation. The operational incident is mapped into the simulation as the incapacitation of the participant to process outgoing payments, ie the inability to submit transactions for the whole day.[18] This assumption is extreme but plausible. As shown in Schmitz and Puhr (2007), shorter outages of participants may lead to payment delays but not to unsettled payments.

In Figure 6.1, the upper panel shows that about 27.5 per cent of all simulations (8604) do not lead to contagion. Another 26.3 per cent (8230) yield one contagious default and 33.1 per cent (10375) two to five, while 13.1 per cent (4102) lead to more than five contagious defaults with a maximum across the 31311 simulations of 33.

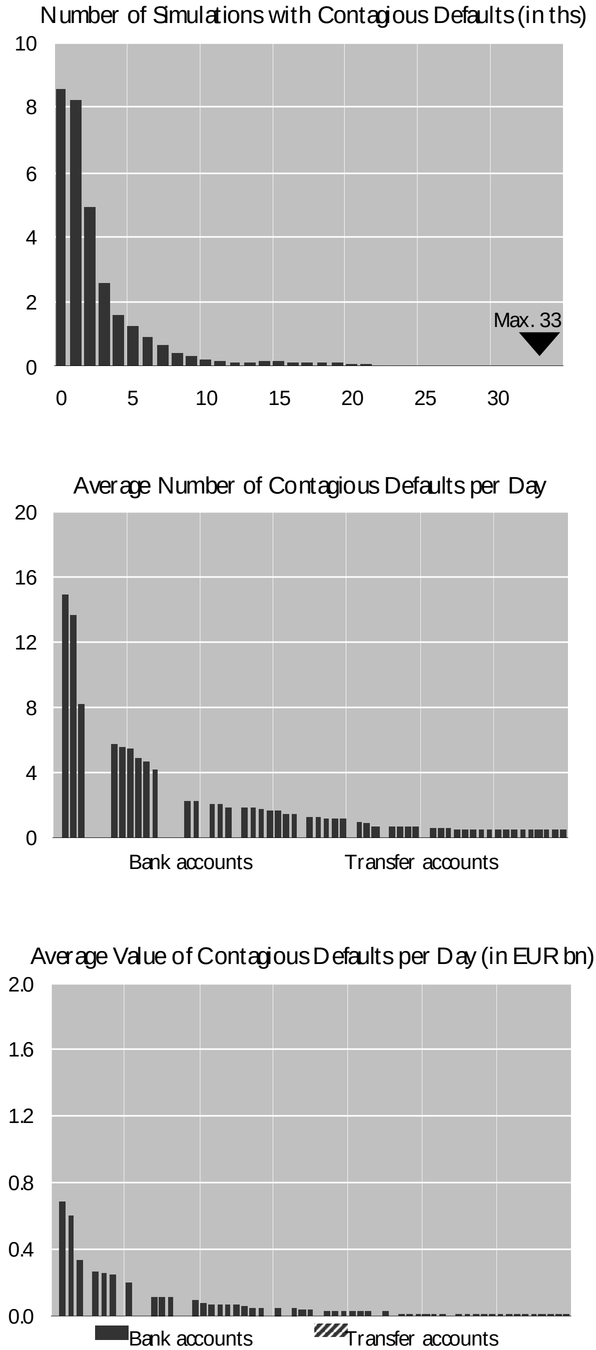
---

[16] For more details on simulations, their motivation, and their design, see Schmitz and Puhr (2007). The operation of ARTIS was discontinued after 16 November 2007, due to the introduction of TARGET2.

[17] Transfer accounts are ARTIS accounts held by other ESCB central banks at OeNB. All national TARGET components are directly linked by transfer accounts. All transactions to and from the respective country and Austria are routed via these transfer accounts.

[18] It is assumed that the resulting illiquidity of the participant is not interpreted as potential insolvency by other participants of the payment system and the financial system at large. In addition, ARTIS provides business continuity arrangements for participants. We tested their impact in Schmitz and Puhr (2007), but disregard them in this paper, as they are of little relevance for the interaction between network topology and contagion.

Figure 6.1 **Contagious defaults in ARTIS**



Source: Own calculations based on daily simulations of the 50 banks and 13 Transfer accounts that formed part of the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays).

The other two panels in Figure 6.1 show the average contagious defaults per simulation (the former in terms of number of participants (banks or transfer accounts) with unsettled payments and the latter in terms of average value of unsettled payments, both due to contagious defaults per simulation). As argued above, we suggest using this information to derive a set of systemically relevant ARTIS participants. If we set the threshold[19] for example in terms of the value of contagious defaults, only participants that cause at least an average value of EUR 48.5 million in unsettled payments (or 0.1 per cent of average value of transactions settled across days), we see the number of systemically relevant participants shrink to 24 (17 banks, seven Transfer accounts). That equals about seven percent of the average of 230 banks in ARTIS (during the sample period) and about two percent of the average of 850 banks in Austria. These results suggest that the supervision of operational risk in banks' payment processing capacity could focus on a relatively small set of systemically relevant banks in Austria and on their business continuity arrangements.

6.4 Univariate analysis of structure and stability

Following the argument in Section 6.2 (choice of structural measures) and in Section 6.3 (choice of stability measures), we provide a selection of univariate results that in turn provide the intuition for our multivariate analysis in Section 6.5. We look at the variation of network indicators at the network-level across days (4.1) and at the node-level across stricken participants (4.2) to explain the variation of contagion across days and across stricken participants.

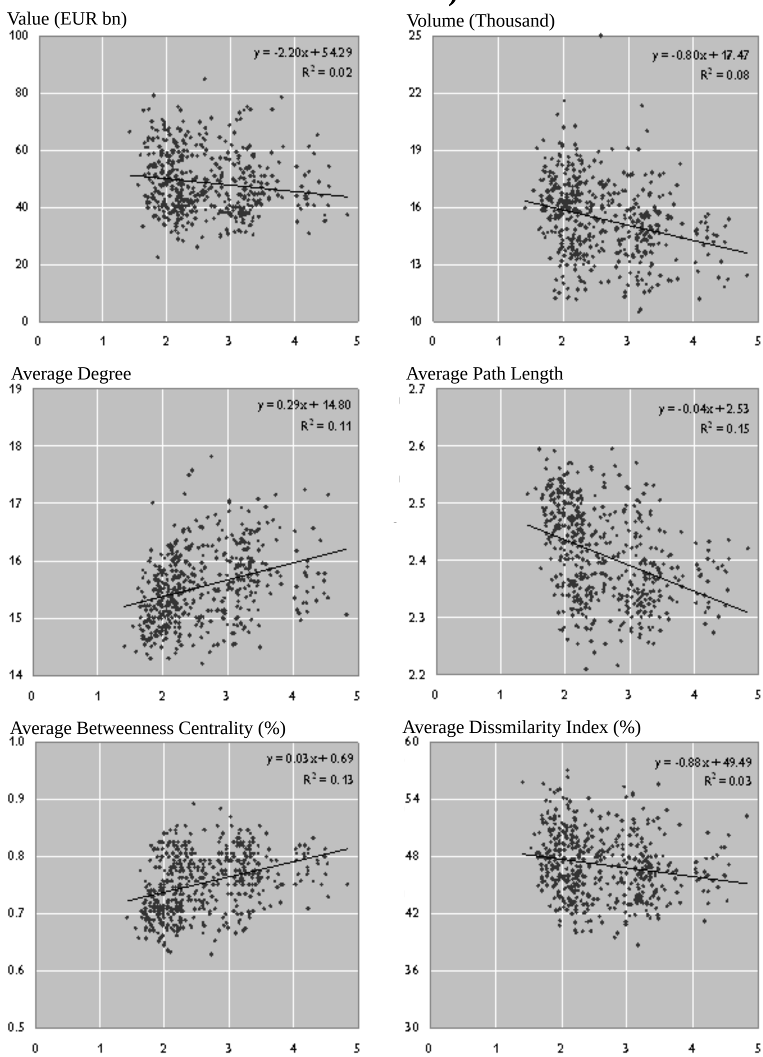
6.4.1 Network-level

In the top two panels of Fiure 6.2 we depict the scatter plot for the value (left hand panel) and the volume of all payments (right hand panel) submitted to ARTIS on the y-axis and the number of contagious defaults in terms of the number of participants with unsettled payments (daily averages across scenarios) per day on the x-axis. The variation of value explains 2 per cent and the variation of

[19] To some extent that threshold is arbitrary and depends on the risk aversion of the supervisory authority.

volume explains 8 per cent of the variation in contagion impact per day.[20]

Figure 6.2 **ARTIS Network Indicators (Network-level) vs Contagion (Daily Average of Number of Participants with Unsettled Payments across Scenarios)**



Source: Own calculations based on daily ARTIS GSCC network indicator averages and on daily simulations of the 50 banks and 13 Transfer accounts that were part of the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays).

[20] Neither volume nor value is significant at the common confidence levels.

The explanatory values of the variables value and volume are low. Do network indicators perform any better? In the four other panels of Figure 6.2 we look at the following indicators (in and out, unweighted, undirected): degree, average path length, betweenness centrality and dissimilarity index. The average path length (15 per cent) and betweenness centrality (13 per cent) have the highest explanatory values. The daily variation in degree explains 10 per cent of the variation in contagion and that of the dissimilarity index explains only 3 per cent. Although the explanatory powers of three of the network indicators are higher than those of value and volume, the levels are still low, and tests of significance fail at all common confidence levels.

The highest explanatory power of any of the remaining 39 indicators is 15.4 per cent (average number-weighted clustering coefficient), while a number of indicators have no explanatory power at all. The univariate analysis suggests that daily variations in network indicators at the network-level across days are of limited use in the stability analysis of ARTIS. However, this does not preclude that either (i) network indicators at the node-level or (ii) structural differences across networks might influence their (relative) resilience.

The study of the latter, for which we currently lack the data, could be subject of future research. Given the fact that other large value payment systems which display considerable differences in size share notable structural commonalities with ARTIS,[21] some doubt is justified as to whether network indicators at the network-level could explain contagious effects in other large value payment systems. That leaves the question whether the different positions of the nodes that experience the operational incident in the network account for this variation.

6.4.2 Node-level

In the two upper panels of Figure 6.3 we plot the value and volume of payments of the stricken node in each simulation against its contagion effect in terms of number of participants with unsettled payments. The variations in value and volume across simulations explain 73 per cent and 68 per cent of the variation of the contagion impact across

[21] As shown eg in the comparison of FedWire and ARTIS in Schmitz and Puhr (2007).

simulations.[22] The slopes have the expected signs: more active nodes cause more contagion.

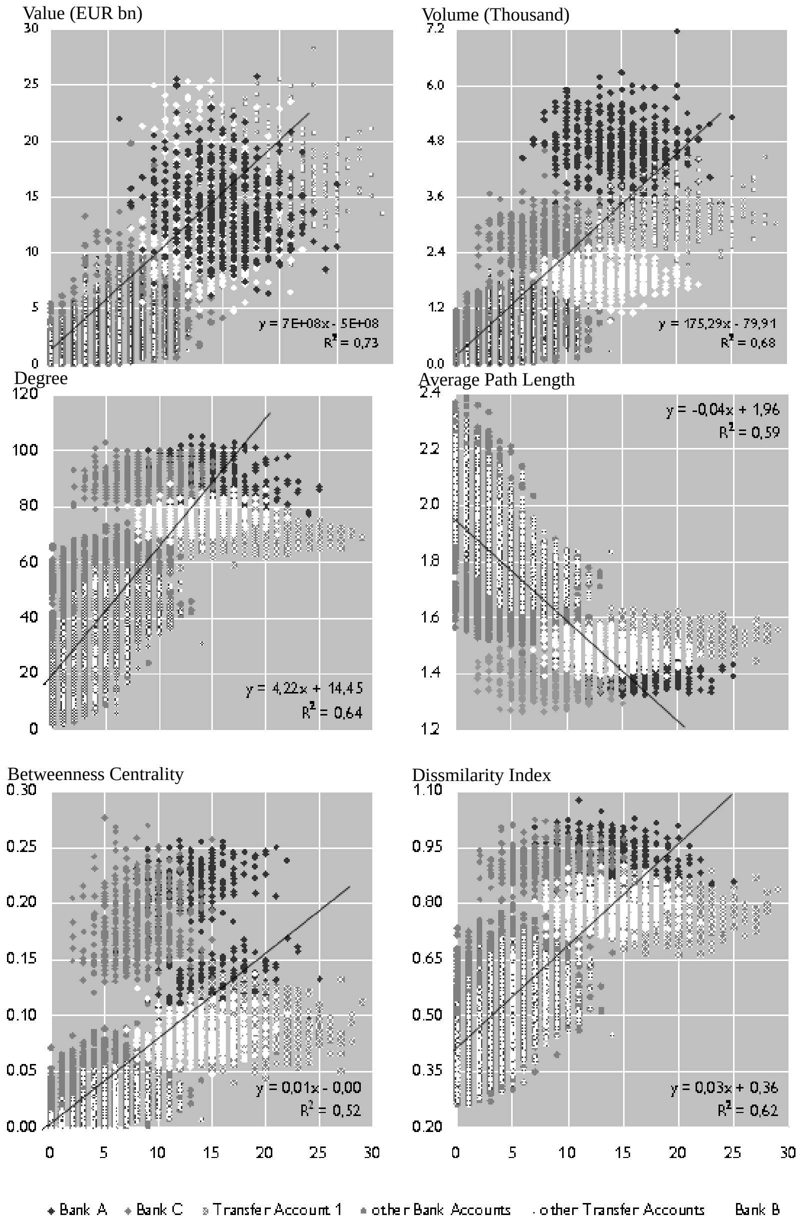
Given the large number of data points (31311) and the variation of the stability measure across ARTIS participants,[23] we differentiate in Figure 6.3 between shocks to banks and Transfer accounts. In addition we highlight the three most active banks and the most active Transfer account. The differentiation reveals a pronounced grouping in both panels. In the right hand panel it also points to structural differences in contagion impact not accounted for by variations in volume. The most active Transfer account and one of the three banks tend to group below the regression line (ie they cause more contagion than estimated by volume of transactions) and the other two banks group above the regression line (ie they cause less contagion than estimated by volume of transactions).

Given the low explanatory values of the variables value and volume at the network-level, we asked whether network indicators at the network-level perform any better and found that they did, albeit at sill at modest levels. Given the already high explanatory power of value and volume at the node-level, we look at whether our four previously selected network indicators at the node-level[24] can again add to that.

We find that the explanatory values of all four network indicators are quite high;[25] the most simple measure degree yields an R^2 of 64 per cent, variations in average path length across simulations account for 59 per cent of the variation in the number of contagious defaults across simulations. The more complex measures betweenness centrality and dissimilarity index yield R^2s of 52 and 62 per cent, respectively. The signs of all slopes are in line with expectations: simulations, in which more active and more central nodes are shocked, feature higher contagion impacts. Moreover, all indicators are highly significant at common confidence levels and also in the order of magnitude of the reported interaction between betweenness centrality and contagious defaults for the Austrian interbank market.[26]

[22] Both volume and value are highly significant at all common confidence levels.
[23] As shown in Section 3, see for instance Figure 3.1.
[24] Previously selected network indicators include: degree, average path length, betweenness centrality and dissimilarity index.
[25] We present the simple linear regression results in order to provide an indication of relative performance and to motivate our approach in the panel data analysis rather than suggesting that OLS is appropriate per se.
[26] See Boss et al (2004).

Figure 6.3 **ARTIS network indicators (network-level) vs contagion (daily average of number of participants with unsettled payments across scenarios)**



Source: Own calculations based on daily node-level ARTIS GSCC network indicators and daily simulations of the 50 banks and 13 Transfer accounts that were part of the ARTIS GSCC every day from 16 November 2005 to 16 November 2007 (excluding Austrian holidays).

The remaining 65 network indicators yield explanatory values between nil (number-weighted average path length based on payments received) and 77 per cent (relative volume of payments received).[27] The results demonstrate that network indicators at the node-level seem to explain large parts of the variation in contagion across stricken participants in a univariate setting. However, they seem to add little to the high explanatory values of the traditional measures of activity (value and volume). Furthermore, the large set of available indicators and the huge differences in their explanatory values pose the problem of data mining.

The aforementioned differentiation according to the stricken ARTIS participant (bank or Transfer account) confirms the pronounced structural differences in contagion impacts not explained by variations in volume. In all four network indicator based panels of Figure 6.3 simulations based on the most active Transfer account cluster on the right hand side of the regression line, while those based on the two aforementioned banks cluster on the left side.[28] This result points to structural differences in contagion impact which are not explained by measures of activity or network indicators and which warrant further research.

We also investigate the interaction between network structure and network stability for other measures of contagion; as an example, we present the value of unsettled payments. As above, we start with an analysis of the explanatory value of node size (value and volume of payments originating at the node). Both values are lower than the respective previous results presented in Figure 6.3. Moreover, only value is significant, explaining 54 per cent of variation in contagion. Albeit the explanatory power of 39 per cent, volume is not significant at common confidence levels.

How well do the network indicators at the node-level fare in comparison? The R^2s of the four previously presented network indicators range between 24 and 29 per cent and are therefore considerably lower than (i) the respective values for the measures of node size above, but (ii) also their respective values when explaining

[27] Due to the large number of observations and the ensuing degrees of freedom, any indicator with an R^2 of 0.51 or higher is significant at all common confidence levels, whereas for indicators with an R^2 of 0.50 or below the null hypothesis cannot be rejected.

[28] The graphs might also be read as suggesting non-linearity; but we prefer the interpretation of structural differences, which we can then exploit in the panel data analysis.

contagion as measured by the number of participants with unsettled payments.[29]

We conclude that if we measure contagion by the value of unsettled payments, network indicators are clearly dominated by the traditional measures of size. Comparing the results for the two measures of contagion, number of participants with unsettled payments versus value of unsettled payments, reveals that contagion under the latter measure is much harder to explain even by the superior traditional variables.

Table 6.2 **Correlations between ARTIS network indicators (node-level)**

	Volume	Value	Avg. PL	Degree	Conn.	Clust.	Btw. C.	Dissim.
Volume	100.0%	89.0%	84.0%	83.0%	-77.0%	-57.0%	89.0%	85.0%
Value		100.0%	76.0%	75.0%	-70.0%	-52.0%	77.0%	78.0%
Avg. PL			100.0%	99.0%	-96.0%	-72.0%	85.0%	95.0%
Degree				100.0%	-97.0%	-72.0%	85.0%	93.0%
Conn.					100.0%	62.0%	-79.0%	-85.0%
Clust.						100.0%	-56.0%	-78.0%
Btw. C.							100.0%	87.0%
Dissim.								100.0%

Source: Own calculations based on daily averages of the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Network indicators include: Average Path Length (Avg. PL), Connectivity (Conn.), Clustering Index (Clust.), Betweenness Centrality (Btw. C.), Dissimilarity Index (Dissim.).

In order to corroborate our findings from the univariate analysis, that network indicators at the node-level do not add much value to stability analysis, we present the correlations between traditional measures of activity (value and volume) and selected network indicators in Table 6.2. The data reveal that particularly indicators of centrality are highly correlated with value and volume. Nevertheless, the question remains open whether these indicators add some explanation in a multivariate setting.

[29] Individual explanatory values are as follows: degree 28 per cent, average path length 25 percent, betweenness centrality 24 per cent and dissimilarity index 29 per cent, none of which are significant at common confidence levels.

6.5 Multivariate analysis of structure and stability

In this section we study the robustness of our findings in the univariate setting of section 6.4 in an exploratory multivariate study. We focus on combining one of the traditional measures of node size (value) with network indicators at the node and at the network-level as well as additional control variables (eg beginning of day liquidity at individual nodes, dummy variable for Transfer accounts) in a panel data setting to answer the following four questions:

1. What explains the variations in contagion across days within scenarios?
2. What explains the variations in contagion across scenarios on each day?
3. Are network indicators at the network and/or at the node-level significant in this context?
4. What is the explanatory contribution of the network indicators at the network and/or at the node-level in the context of questions 1 and 2?

In Section 6.5.1 we introduce our measures of contagion as dependent variables. We try to explain their variation with the three groups of independent variables discussed in Section 6.5.2: first, the independent variables at the network-level, which are constant across panels but vary over time; second, the independent variables at the node-level that vary over both time and scenarios; third, we add a dummy variable for Transfer accounts to corroborate the findings of some of the hitherto unexplained structural particularities discovered in the scatter plots of Section 6.4. In Section 6.5.3 we introduce our model as well as its assumptions and the estimation method. In Section 6.5.4 we present the results.

6.5.1 Dependent variables

As dependent variables we focus on our three measures of contagion (excluding the impact on the stricken bank). Three different measures of contagion provide a unique opportunity to check the robustness of the models and the parameter estimates. Table 6.3 shows that the means and standard deviations of the dependent variables differ substantially over time and across scenarios. First, the number of participants with unsettled payments amounts to an overall daily average of 2.6 (overall standard deviation 3.8) with a minimum of 0 and a maximum of 33. The variation between scenarios (between standard deviation 3.5) is much higher than the variation over time within scenarios (within standard deviation 1.4). Second, the average daily volume of unsettled payments is 7.6 per day (overall standard deviation 21.7). The lowest value is 0 the highest value 1 172. In this case the standard deviation between scenarios (14) is slightly lower than the one within scenarios (16.7). Third, the daily value of unsettled payments averages EUR 112 million with a range of 0 to 10.7 billion. The overall standard deviation is EUR 335 million and the between standard deviation is much higher (EUR 284 million) than the within one (EUR 181 million).

Table 6.3 **Dependent variables (measures of contagion)**

Variable		Mean	Std.Dev.	Min.	Max.	Obs.
Number of participants with unsettled payments	overall between within	2.6	3.8 3.5 1.4	0	33	N=31311 n=63 T=497
Volume of unsettled payments	overall between within	7.6	21.7 14 16.7	0	1172	N=31311 n=63 T=497
Value of unsettled payments (in EUR billion)	overall between within	0.11	0.34 0.28 0.18	0	10.7 3	N=31311 n=63 T=497

Source: Own calculations based on the ARTIS GSCC data from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Standard Deviation (Std.Dev.), Observations (Obs.).

6.5.2 Independent variables

We split the independent variables into two groups, network-level and node-level.

First, the independent variables at the network-level are constant across panels but vary over time ([Z] in the model below): They include aggregate liquidity (Liquidity[30]), a traditional measure of network size (aggregate value of all transactions (Value (Network)), and a range of network indicators at the network-level.

[30] It corresponds to our traditional measure of size value in the univariate analysis. We chose to rename here to facilitate economic interpretation and intuition.

Table 6.4 displays the independent variables at the network level – Liquidity, Value(Network), and the network indicators at the network-level (including mean and standard deviation for all variables). We define aggregate liquidity (Liquidity) in the system (mean: EUR 18.3 billion, standard deviation: EUR 3.2 billion, Table 6.4) as the sum of beginning of day balances (EUR 7.5 billion; EUR 0.8 billion) and unencumbered collateral[31] (EUR 10.8 billion; EUR 2.9 billion) across participants in the system.[32]

Turning to the network indicators at the network level, it is apparent that the relative standard deviations of the network indicators across days – often only small fractions of the respective means – are much lower than those of the measures of contagion (exceeding the respective means).[33]

[31] We simply aggregate across beginning of day balances and unencumbered collateral because the latter can be liquidised via interest free Daylight Overdrafts at OeNB within minutes (for details see Schmitz and Puhr, 2007).

[32] Focusing on real historical data might restrict the generalisation of the results to other payment systems in which participants might follow different liquidity policies. Eg in systems, that experience frequent operational outages, it might be rational for participants to hold sufficient liquidity to settle all outgoing payments. As a consequence there would be no contagion.

[33] While standard deviation is not the ideal parameter to describe the distributions of the network indicator, they are helpful in pointing out the differences between within and between scenario variation in our data set.

Table 6.4 **Independent variables at the network level: liquidity, netvalue, and network indicators at network-level**

Variable		Mean	Std.Dev.	Min.	Max.	Obs.
Liquidity (in EUR billion)	overall	18.27	3.23	11.68	32.37	N=31311
	between		0.00			n=63
	within		3.23			T=497
Value (Network) (in EUR billion)	overall	48.90	10.62	22.88	85.60	N=31311
	between		0.00			n=63
	within		10.62			T=497
Average Degree (Network)	overall	12.3	0.4	11.3	14.4	N=31311
	between		0.0			n=63
	within		0.4			T=497
Average Connectivity (Network)	overall	0.040	0.003	0.030	0.050	N=31311
	between		0.000			n=63
	within		0.003			T=497
Average Path Length (Network)	overall	2.50	0.06	2.40	2.70	N=31311
	between		0.00			n=63
	within		0.06			T=497
Average Cluster Index (Network)	overall	0.40	0.03	0.40	0.50	N=31311
	between		0.00			n=63
	within		0.03			T=497
Average Betweeness Centrality (Network)	overall	0.0050	0.0003	0.0040	0.0060	N=31311
	between		0.0000			n=63
	within		0.0003			T=497
Average Dissimilarity Index (Network)	overall	1.3	0.9	0.6	5.2	N=31311
	between		0.0			n=63
	within		0.9			T=497

Source: Own calculations based on ARTIS data for the independent variables from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Standard Deviation (Std.Dev.), Observations (Obs.). N.B. Data differs from that in Table 6.1 (network indicators for all accounts in the GSCC).

Second, the independent variables at the node-level (Liquidity loss and the network indicators at the node level; Table 6.5 plus a dummy variable for Transfer accounts) vary, both over time and across scenarios ([X] in the model below). Liquidity Loss is measured as the value of payments that were due by the stricken bank but were not processed in the simulations due to operational problems at the stricken bank.[34] The dummy variable D for Transfer accounts took the values 0 or 1 and entered the models as D×Liquidity Loss to measure the deviation of the impact of the Liquidity Loss variable in the case of Transfer accounts from the average across all participants.

[34] We have also used alternative proxies for the impact of the operational problem of the stricken bank on liquidity in the system, such as liquidity drain and liquidity sink, which yield similar results both in terms of sign and significance.

Table 6.5 **Independent variables at the node level: liquidity loss and network indicators at node-level**

Variable		Mean	Std.Dev.	Min.	Max.	Obs.
Liquidity Loss	overall	0.76	1.69	0.00	22.05	N=31311
(in EUR billion)	between					n=63
	within					T=497
Average	overall	25.4	19.9	2	105	N=31311
Degree (Node)	between		19.8			n=63
	within		3			T=497
Average	overall	0.200	0.15	0.015	0.800	N=31311
Connectivity (Node)	between		0.150			n=63
	within		0.02			T=497
Average	overall	1.90	0.18	1.20	2.30	N=31311
Path Length (Node)	between		0.18			n=63
	within		0.02			T=497
Average	overall	0.50	0.2	0.13	1.00	N=31311
Cluster Index (Node)	between		0.19			n=63
	within		0.08			T=497
Average Betweeness	overall	0.1360	0.0349	0.0000	0.2760	N=31311
Centrality (Node)	between		0.0341			n=63
	within		0.0086			T=497
Average Dissimilarity	overall	0.44	0.13	0.26	1.07	N=31311
Index (Node)	between		0.1			n=63
	within		0.03			T=497

Source: Own calculations based on ARTIS data for the 63 accounts defining the scenarios from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Standard Deviation (Std.Dev.), Observations (Obs.).

Table 6.5 also shows that the standard deviations of the network indicators at the node-level are much higher across scenarios than within scenarios over time. Similarly, their means and standard deviations are much higher than those of the network indicators at the network-level; eg the mean of the average degree at the network-level across days is 12.4 (standard deviation 0.4) while the mean of the degree at the node-level across scenarios and across days is 25.4 (standard deviation 19.9).

6.5.3 Models, specifications, and estimation

We estimate the following static fixed effects model

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_{63} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ X_{63} \end{bmatrix} \beta_I + [Z]\beta_{II} + \begin{bmatrix} \nu_1 \\ \nu_2 \\ \cdot \\ \nu_{63} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_{63} \end{bmatrix}$$

The dependent variables y_1 to y_{63} are 497×1 vectors containing daily values for the dependent variable. The vector $[y_1 \dots y_{63}]$ therefore has 31311 elements. In theory, the matrices X_1 to X_{63} would be 497×9 dimensional, as they contain daily observations of the independent variables Liquidity Loss, D×Liquidity Loss, and the six network indicators at the node-level plus the constant term. However, in practice the six network indicators are highly correlated. Similarly, the vector Z would contain Liquidity and the network indicators at the network-level and would have 497×7 dimensions. But the network indicators at the node-level are also highly correlated (see Table 6.2), which could lead to multicollinearity.

The vectors of parameters β_I and β_{II} are to be estimated. The 63×1 dimensional vector $[\nu_1 \dots \nu_{63}]$ contains the scenario specific unobservable time-invariant regressors and the 497×1 dimensional vector $[\varepsilon_1 \dots \varepsilon_{63}]$ consists of the standard error term for each observation (31311 elements). Our panel is balanced, as we conduct simulations for all scenarios in all periods and the number of simulations is equal for all days in the sample period.

In order to avoid multicolinearity, we estimated a basic model for each of the measures of contagion consisting of the independent variables Liquidity, Liquidity Loss and D×Liquidity (specification 1 in Tables 6.6, 6.7, and 6.8). The choice of variables for this basic model rests on previous results and economic intuition, which indicate that i) the aggregate liquidity of the system (Liquidity) reduces contagion[35], ii) Liquidity Loss is significantly positively correlated with contagion in the univariate analysis, and iii) Transfer accounts display interesting particularities in the univariate analysis, which warrant further attention. To this model structure we add the variable Value (Network) (specification 2) or a particular network indicator at the network and at the node-level (specifications 3 to 8).[36]

The following equation is specification 5 of model 1 and explains the variations in the number of participants with unsettled payments across days and scenarios by a constant and the standard ingredients Liquidity, Liquidity Loss and D×Liquidity Loss plus the average path length at the node-level and the network-level.

[35] See the papers in Leinonen (2005).

[36] Why does specification 2 only contain the network level variable and not the node level variable? Remember that the corresponding node level variable is already contained in the specification as Liquidity Loss.

$$\text{Numberofparticipants}_{it} = \beta_1 + \beta_2 \text{Liquidity}_t + \beta_3 \text{LiquidityLoss}_{it} + \beta_4 D \times \text{LiquidityLoss}_{it} + \beta_5 \text{AvgPathLength(Node)}_{it} + \beta_6 \text{AvgPathLength(Network)}_t + \nu_i + \varepsilon_{it}$$

Similarly specification 5 of model 2 would explain the variations of the number of unsettled payments across days and scenarios with the same independent variables

$$\text{Numberofunsettledpayments}_{it} = \beta_1 + \beta_2 \text{Liquidity}_t + \beta_3 \text{LiquidityLoss}_{it} + \beta_4 D \times \text{LiquidityLoss}_{it} + \beta_5 \text{AvgPathLength(Node)}_{it} + \beta_6 \text{AvgPathLength(Network)}_t + \nu_i + \varepsilon_{it}$$

Similarly specification 5 of model 3 would contain the variations of the value of unsettled payments across days as dependent variable and the same independent variables.

$$\text{Valueofunsettledpayments}_{it} = \beta_1 + \beta_2 \text{Liquidity}_t + \beta_3 \text{LiquidityLoss}_{it} + \beta_4 D \times \text{LiquidityLoss}_{it} + \beta_5 \text{AvgPathLength(Node)}_{it} + \beta_6 \text{AvgPathLength(Network)}_t + \nu_i + \varepsilon_{it}$$

There are some basic assumptions regarding static fixed effects models. One states that the error term ε is uncorrelated with past, present, and future values of the independent variables. This assumption of strict exogeneity ensures that the agents whose behaviour is modelled are not influenced by past realisations of the error term ε. Moreover, cross-panel and cross-time conditional homoskedasticity means that the conditional variance of the error term ε – given the time-invariant unobservable scenario specific effect ν – is constant across scenarios and across time. Furthermore serial independence presupposes that the error terms are serially independent within panels and cross-panel independence that they are independent across scenarios.

Do these assumptions hold in our dataset? The first one holds by virtue of the simulation design, since we do not model banks' behavioural reactions to operational shocks. The values of explanatory variables are historic observations of a world without operational shocks and hence without contagion. Consequently, there are no observable, unexplained variations in contagion – past realisations of error terms – which could influence banks' behaviour: eg banks cannot adjust their liquidity reserves or payment behaviour to account for error terms which are the results of counterfactual simulations. The other three assumptions are not fulfilled.[37] Hence an ordinary least squares (OLS) estimate of a standard fixed-effects model would yield inconsistent and biased standard errors. We employ an alternative estimator that accounts for conditional heteroskedasticity, serial correlation, and cross-sectional dependence of the error terms ε.

We apply a panel-corrected standard error estimator with panel specific autocorrelations where the parameters are estimated by Prais-Winsten regression. The parameter estimates are conditional on the estimates of autocorrelation parameters in each panel. The estimator uses a feasible generalised least squares (FGLS) estimate of the variance-covariance matrix which is asymptotically efficient under the assumed covariance structure of the disturbance terms (heteroskedasticity and contemporaneous correlation across panels).

The 63 panels produce 63 variance estimates and 1953 covariance estimates. Together with the 63 autocorrelation estimates and (up to) 6 parameters, a total of 2085 parameter estimates are required. The estimation procedure yields consistent standard errors - at the expense of a large loss in degrees of freedom (2 079). However, with 31311 observations, the degrees of freedom are still large.

Beck and Katz (1995) argue that full FGLS estimates are overly optimistic and that the Prais-Winsten estimator is superior. Although they derived their results for data comprising 10 to 20 panels and 10 to 40 time periods, we also preferred the Prais-Winsten estimator for our larger data set, mainly because the large number of observations in each panel (497) supports the asymptotic behaviour of the panel-specific autocorrelations.

6.5.4 Results of Model 1 (number of participants with unsettled payments)

We present the results of the panel data estimates in three tables (Table 6.6, 6.7, and 6.8), one for each measure of contagion. Table 6.6 summarises the results for the 8 specifications of Model 1. The dependent variable is the number of participants with unsettled payments, ie the number of banks (or transfer accounts; excluding the stricken bank/transfer accounts) with unsettled payments at the end of the day.

[37] See Annex – Test results for the static fixed effects model.

Table 6.6 **Results Model 1 (number of participants with unsettled payments)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	3.28	4.02	3.38	2.32	14.73	6.97	4.63	-0.93
	13.71***	16.5***	3.8***	5.14***	10.1***	14.84***	7.25***	-3.85***
Liquidity	-0.10	-0.10	-0.10	-0.08	-0.09	-0.10	-0.10	-0.12
	-8.19***	-8.05***	-9.16***	-7.87***	-7.56***	-16***	-7.67***	-9.65***
Liquidity Loss	1.65	1.66	0.78	0.81	1.01	1.43	1.07	0.94
	59.21***	58.92***	28.64***	30.8***	38.61***	55.18***	33.78***	33.17***
D x Liquidity Loss	0.25	0.25	0.61	0.60	0.53	0.26	0.54	0.51
	8.99***	8.93***	23.28***	23.92***	20.57***	9.66***	19.29***	19.42***
Degree (Node)			0.09					
			59.92***					
Connectivity (Node)				12.23				
				59.88***				
Avg. Path Length (Node)					-8.09			
					-56.71***			
Cluter Index (Node)						-4.00		
						-51.65***		
Betw. Cenrality (Node)							33.80	
							26.37***	
Dissimilarity Index (Node)								11.13
								40.07***
Value (Network)		-0.02						
		-7.76***						
Degree (Network)			-0.16					
			-2.24***					
Connectivity (Network)				-32.51				
				-3.25***				
Avg. Path Length (Network)					1.51			
					2.52***			
Cluter Index (Network)						-3.60		
						-3.42***		
Betw. Cenrality (Network)							-342.26	
							-3.00***	
Dissimilarity Index								0.03
(Network)								0.82
R^2	69.23	69.35	69.96	72.41	71.58	70.88	66.54	68.69
Relative Impact of Transfer Account (in %)	15%	15%	78%	74%	52%	18%	50%	54%

Source: Own calculations based on daily network- and node-level ARTIS GSCC indicators and on daily simulations of the 50 banks and 13 Transfer accounts in the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Numbers (1) to (8) provide results for the respective model specifications, each including the independent variables for which results are provided. These values are parameter estimates (upper) and corresponding z-values (lower). * denotes significance at the 90 per cent confidence level, ** at 95 and *** at 99 per cent.

First we estimated specification 1 – column (1) – as the most parsimonious model with only three explanatory variables: Liquidity, Liquidity Loss, and the dummy variable (D×Liquidity Loss). All three variables are highly significant and have the expected signs: higher values of aggregate liquidity in the system reduce contagion; large values of liquidity loss at the stricken bank increase contagion; and the contagion impact of operational shocks at Transfer accounts is significantly higher than that of the average participant in the system (as suggested by Figure 6.3 above).

These results are not particularly surprising, but it is reassuring for our approach that they are robust across all specifications. The goodness of fit of specification 1 is high with an R^2 of 69.23.[38] The relative impact of a stricken Transfer account is 15 per cent higher than the average across stricken banks and Transfer accounts. The parameter values of the three explanatory variables are very robust across specifications. They remain highly significant in all specifications.

Secondly, we add further explanatory variables that capture either network or node characteristics. Since the various network indicators are highly correlated, we add only one indicator in each specification, for both node- and the network-level, in Specifications 3 to 8. In Specification 2 we include a traditional measure of network size (Value of all transactions settled on a specific day) to contrast the explanatory impact of this traditional measure with more sophisticated network indicators (degree, average path length etc.). A higher transactions value in the network is associated with lower contagion. However, the additional variable only has a very small impact on the model's explanatory power (R^2 increases by 0.12 percentage point).

The network indicators at node-level are highly significant. Operational shocks at nodes with higher degree, higher connectivity, higher betweenness centrality or higher dissimilarity indices cause higher contagion. Similarly, nodes with higher average path length and higher cluster indices feature lower contagion. In sum, more connected, more central nodes, and nodes with less mutually connected neighbours cause relatively more contagion, even compared to i) nodes which cause similar liquidity losses, ii) nodes that are either also banks / Transfer accounts, and iii) nodes that experience operational shocks on days with the same level of aggregate liquidity. We conclude that in Model 1 the position of the stricken bank in the network indeed has an impact on contagion.

The network indicators at network-level are significant as well (except for the average dissimilarity index). Lower average degree, average connectivity, average cluster index and average betweenness centrality are associated with a higher contagious impact of an operational problem at a given bank / Transfer account across days. Higher average path length implies a higher contagious impact; ie the denser the network on a specific day, the lower the contagious impact

[38] The R^2s reported in the panel data analysis differs from the OLS R^2. While they are still a useful measure of the model and its specification, they are not equal to the fraction of variation of the dependent variable explained by the estimated equation. They can be interpreted as squared correlations between the estimated and observed dependent variable.

on that day in Model 1. Hence, the structure of the network on a given day in the sample significantly influences contagious.

The additional pairs of explanatory variables do not seem to improve the goodness of fit of the model, though it is somewhat higher than for the simple measure of network activity Value (Network). Adding connectivity at node-level and network-level – specification (4) – increases R^2 from 69.23 to 72.41 per cent. In this respect, the multivariate analysis indeed contradicts the univariate analysis in section 4 above.

The relative impact of Transfer accounts varies greatly across specifications, from roughly 15 per cent – specifications (1), (2), and (6) – to almost 80 per cent in specification (3). However, the relative impact is positive across all specifications and confirms the observation in Figure 6.3 above.

6.5.5 Results of Model 2 (number of unsettled payments)

Table 6.7 summarises the results for the eight specifications of Model 2. The dependent variable is the number of unsettled payments at the end of the day (excluding those of the stricken bank).

In specification (1) the variables Liquidity, Liquidity Loss, and the dummy variable (D×Liquidity Loss) are highly significant and carry the expected signs. Again the parameters and z-values are robust across all specifications. Contagion is lower on days with higher aggregate liquidity. It is higher when the stricken bank planned to transact a higher value of payments on the day of the simulated operational problem. Operational shocks to Transfer accounts have a higher contagion impact than those to bank accounts. The relative impact is more than twice as high in specification (1) of Model 2 (36 per cent) as in the comparable specification of Model 1 (15 per cent). The goodness of fit of the specification is lower in Model 2 (39.79 per cent) than in Model 1 (69.23 per cent). The variation in number of unsettled payments across days and across scenarios is harder to capture than that of the number of participants affected by contagion.

Table 6.7 **Results Model 2 (number of unsettled payments)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	7.32	9.90	-3.06	0.57	35.54	12.55	4.73	-0.36
	7.95***	10.67***	-0.93	0.34	6.38***	6.93***	2.13**	-0.35
Liquidity	-0.28	-0.25	-0.27	-0.20	-0.22	-0.25	-0.26	-0.29
	-5.65***	-5.44***	-6.18***	-4.57***	-4.75***	-5.43***	-5.29***	-6.24***
Liquidity Loss	6.61	6.67	4.97	4.94	5.32	6.29	5.60	5.40
	41.91***	41.84***	24.31***	24.17***	28.22***	38.61***	22.59***	26.62***
D x Liquidity Loss	2.40	2.38	3.10	3.10	2.94	2.45	2.91	2.84
	8.36***	8.29***	10.53***	10.55***	10.08***	8.49***	9.52***	9.72***
Degree (Node)			0.18					
			13.51***					
Connectivity (Node)				24.59				
				13.85***				
Avg. Path Length (Node)					-17.09			
					-13.92***			
Cluter Index (Node)						-6.25		
						-10.84***		
Betw. Cenrality (Node)							53.96	
							5.65***	
Dissimilarity Index (Node)								19.00
								10.02***
Value (Network)		-0.07						
		-8.15***						
Degree (Network)			0.51					
			1.93**					
Connectivity (Network)				30.93				
				0.86				
Avg. Path Length (Network)					1.21			
					0.56			
Cluter Index (Network)						-5.29		
						-1.4*		
Betw. Cenrality (Network)							426.79	
							1.09	
Dissimilarity Index (Network)								0.05
								0.52
R^2	39.79	39.81	39.90	40.01	40.02	39.77	39.72	39.67
Relative Impact of Transfer Account (in %)	36%	36%	62%	63%	55%	39%	52%	53%

Source: Own calculations based on daily network- and node-level ARTIS GSCC indicators and on daily simulations of the 50 banks and 13 Transfer accounts in the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Numbers (1) to (8) provide results for the respective model specifications, each including the independent variables for which results are provided. These values are parameter estimates (upper) and corresponding z-values (lower). * denotes significance at the 90 per cent confidence level, ** at 95 and *** at 99 per cent.

Similar to Model 1 all network indicators at node-level are highly significant and have the same signs as in the respective specifications of Model 1. A participant with higher node degree, connectivity, betweenness centrality and dissimilarity index, but lower average path length and cluster index causes a larger contagion effect than participants with similar values of out-going payments on that day. The position of the stricken node within the network has a significant influence on the contagion caused by an operational shock, even after controlling for aggregate liquidity, for liquidity loss due to operational problems at the stricken bank, and for whether it is a bank or a Transfer account.

Turning to the network indicators at network-level provides the following picture: only Value (Network) is significant at the 99 per cent confidence level. Again, operational problems cause less contagion on days with more network activity. The degree is significant at the 95 per cent confidence level, ie a more connected network is subject to more contagion. This contradicts the finding in Model 1 and also the following finding in Model 2: The clustering index is significant at the 90 per cent level, which implies that a network experiences less contagion if it consists of nodes with more mutually connected neighbours. The other more sophisticated network indicators are not significant in Model 2.

The explanatory value of the additional explanatory variables in specifications (2) to (8) is very low. It increases from 39.79 per cent in specification (1) to at most 40.02 in specification (5). We conclude that network indicators (at both node and network-level) provide little explanation for variations in the number of unsettled payments across days and across scenarios.

The relative impact of an operational shock at a Transfer account varies substantially across specifications, with a minimum of 36 and a maximum of 63 per cent. The minimum is higher, but the maximum is lower in Model 2 than in Model 1. Nevertheless, the dummy is significant in all specifications.

6.5.6 Results of Model 3 (value of unsettled payments)

Table 6.8 summarises the results for the 8 specifications of Model 3. The dependent variable is the value of unsettled payments at the end of the day (excluding payments by the stricken bank).

Specification (1) represents the basic Model 3 with the dependent variable value of unsettled payments and independent variables Liquidity, Liquidity Loss, and the dummy variable (D×Liquidity Loss). All three latter variables are highly significant and have the expected signs. High aggregate liquidity in the system cushions the contagious effect of an operational shock. It increases with the value of payments that the stricken bank would have transferred under business as usual. Operational problems at Transfer accounts cause significantly more contagion than the average across accounts. The relative impact (166 per cent), however, is much larger than in Model 1 (15 per cent) and Model 2 (36 per cent). Unlike in the other two models, the relative impact is quite stable across specifications (1) to (8), ranging from 163 to 174 per cent. The goodness of fit of specification (1) is high, with an R^2 of 70.62.

Table 6.8 **Results Model 3 (value of unsettled payments)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	5.22E-02	7.37E-02	2.47E-02	3.63E-02	2.87E-01	5.94E-02	4.24E-02	5.35E-02
	5.3***	7.16***	0.59	1.7**	4.08***	2.63***	1.44*	4.59***
Liquidity	-2.66W-03	-2.19E-03	-2.70E-03	-2.50E-03	-2.08E-03	-2.68E-03	-2.63E-03	-2.72E-03
	-5.01***	-4.29***	-5.08***	-4.72***	-3.57***	-4.77***	-4.8***	-4.97***
Liquidity Loss	9.56E-02	9.60E-02	9.22E-02	9.04E-02	9.09E-02	9.72E-02	9.65E-02	9.57E-02
	47.6***	47.77***	35.69***	35.17***	38.17***	45.99***	33.87***	36.91***
D x Liquidity Loss	1.59E-01	1.59E-01	1.60E-01	1.61E-01	1.61E-01	1.58E-01	1.58E-01	1.59E-01
	34.54***	34.48***	34.24***	34.39***	34.55***	34.6***	33.08***	34.04***
Degree (Node)			3.33E-04					
			2.34**					
Connectivity (Node)				6.84E-02				
				3.63***				
Avg. Path Length (Node)					-5.76E-02			
					-4.42***			
Cluter Index (Node)						2.78E-02		
						3.66***		
Betw. Cenrality (Node)							-4.65E-02	
							-0.53	
Dissimilarity Index (Node)								-2.52E-03
								-0.12
Value (Network)		-6.24E-04						
		-5.49***						
Degree (Network)			1.76E-03					
			0.52					
Connectivity (Network)				7.28E-02				
				0.16				
Avg. Path Length (Network)					-5.32E-02			
					-1.92**			
Cluter Index (Network)						-5.39E-02		
						-1.11		
Betw. Cenrality (Network)							1.93E+00	
							0.36	
Dissimilarity Index (Network)								5.25E-04
								0.41
R^2	70.62	70.63	70.64	70.65	70.69	70.68	70.60	70.61
Relative Impact of Transfer Account (in %)	166%	166%	174%	178%	177%	163%	164%	166%

Source: Own calculations based on daily network- and node-level ARTIS GSCC indicators and on daily simulations of the 50 banks and 13 Transfer accounts in the ARTIS GSCC from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Numbers (1) to (8) provide results for the respective model specifications, each including the independent variables for which results are provided. These values are parameter estimates (upper) and corresponding z-values (lower). * denotes significance at the 90 per cent confidence level, ** at 95 and *** at 99 per cent.

Four of the six network indicators at node-level are significant; the most sophisticated (betweenness centrality and dissimilarity index) are not. The stricken banks' / Transfer accounts' degree, connectivity, and average path length have the same signs as in the respective specifications in Models 1 and 2. Again we conclude that more central and more connected nodes cause more contagion. However, the cluster index is significant again but changes sign relative to two previous models, ie nodes with higher fractions of mutually connected nodes cause more contagion.

With the exception of the traditional measure Value (Network) and average path length, network indicators at network-level are not significant. Those two, however, have the same signs as in the respective specifications in Models 1 (and 2). On days with a higher total transactions value and/or more dispersed network structure, contagion is lower. However, the more sophisticated network indicators at node and network-level seem to have little bearing on the goodness of fit. R^2 ranges from 70.60 in specification (7) to 70.69 in specification (5). The most parsimonious specification features an R^2 of 70.62.

6.5.7 Overall results of the multivariate analysis

The explanatory value of all three models is high for all eight specifications. It ranges from about 40 per cent for Model 2 to roughly 70 per cent for Models 1 and 3. In the first case, the variation within scenarios is higher than between scenarios, while in the other two cases the opposite is true. Based on the higher R^2 values for Models 1 and 3, we conclude that our models are better suited to explain variation of the contagion impact between scenarios than over time within scenarios.

The three variables aggregate Liquidity, Liquidity Loss, and the dummy variable for Transfer accounts (D×Liquidity Loss) are highly significant across all models and across all specifications. They have the same sign in all cases. We regard the following results as robust

- The contagion effect is lower on days with more aggregate liquidity in the system.[39]
- The contagion effect is lower for scenarios and days with lower liquidity loss due to operational shocks.
- The system is significantly more vulnerable to operational shocks that hit Transfer accounts. These have special characteristics which are neither fully captured by the Liquidity Loss measure nor by the position within the network.

Three network indicators at node-level are significant and have the same signs in Models 1 to 3: degree (+), connectivity (+), and average path length (–). We conclude that operational shocks at more connected and more central nodes cause more contagion, even after controlling for variations in liquidity loss (which can also be regarded as an indicator of their importance/activity in the payment system) and for whether they are Transfer accounts or banks.

[39] We re-estimated specifications (1) to (8) of Models 1 to 3 in sub-periods of the sample period. In one sub-period, *Liquidity* was not significant in some specifications.

None of the more sophisticated network indicators at network-level are significant in any of the three models. Only the traditional measure of network size (Value (Network), the aggregate value of payments in the network on a given day) is significant across models and features the same sign. Days with higher transaction activity are associated with lower contagion, even after controlling for aggregate liquidity and liquidity loss, and for whether the shock hits a Transfer account or a bank.

The additional explanatory power of network indicators at node- and network-level seems to be very limited. The liquidity situation of the system, the liquidity loss due to operational shock, and the hitherto unknown special characteristics of Transfer accounts can explain the variation in contagion across days and across scenarios already very well. The position of the stricken bank / Transfer account in the network and the structure of the network on the specific day of the shock add little explanatory value.

In further research we will focus on two main issues: first, the impact of different liquidity strategies of banks; in our approach aggregate liquidity conceals the potential impact of the distribution of liquidity in the system. Furthermore, we could run simulations on different levels of theoretical liquidity at individual banks and test for the impact of liquidity at individual-bank level. Second, we find that network indicators at network level add little to a stability analysis within a given network. However, combining simulation data from different networks might show that network indicators play a more prominent role in stability analysis between networks than within networks.

6.6 Summary

The analysis of the network indicators of ARTIS shows that the network is compact. This is mostly due to the fact that almost all active nodes are linked to a small number of nodes at the centre of the network (the largest banks and the most active Transfer accounts). This network structure is quite stable across days.

We conducted 31311 simulations based on 63 different scenarios for 497 transaction days from 16 November 2005 to 16 November 2007 (excluding Austrian holidays). Although the scenarios focus solely on banks and Transfer accounts in the GSCC on all days, more than a quarter of all simulations do not lead to contagion (in terms of the number of banks with unsettled payments), and two-fifths yield only one or two contagious defaults. We arbitrarily define a conservative threshold for the systemic importance of an account based on the average contagion it causes across day. An account is deemed systemically important if it causes an average contagious impact of at least EUR 48.5 million in unsettled payments (0.1 per cent of average transactions settled across days). We find that only a very small number of participants are systemically important (seven per cent of banks in the network and about 50 per cent of Transfer accounts). The simulation results suggest that the ARTIS system is remarkably stable with respect to an operational incident at a single participant. The strong contagion impact of the Transfer accounts is an interesting feature revealed by the simulations and suggests that the removal of Transfer accounts by the single shared platform in TARGET 2 can improve resilience relative to the old TARGET system.

The simulation results reveal that contagion varies substantially across scenarios and across days. In order to explore the determinants of variation, we employ a panel data analysis. We test eight specifications of three models (based on three different measures of contagion). Specification (1) in each model is the most parsimonious one with three independent variables: Liquidity (aggregate liquidity in the system on any given day), Liquidity Loss (the value of payments due by the stricken account on any given day), and a dummy variable (D×Liquidity Loss) for the Transfer accounts in the panel. We find that the contagion effect is lower on days with more aggregate liquidity in the system as well as in scenarios and on days with lower liquidity loss due to operational shocks. Operational shocks at Transfer accounts render the network significantly more vulnerable to operational shocks.

Over recent years, payment system research has increasingly focused on network analysis. We apply our very rich data set to empirically test the interaction between network structure and network stability for the first time. Specifications (2) to (8) extend the basic model by including network indicators at node- and network level. The results for the network indicators at node level suggest that operational shocks at more connected and more central nodes cause more contagion. The results for the network indicators demonstrate that operational shocks on days with higher transaction activity cause less contagion. These results are highly robust across models and across specifications. But the more sophisticated network indicators at the network level are insignificant.

Furthermore, we find that the most parsimonious specification (1) features high goodness of fit for Models 1 and 3 (dependent variables number of participants with unsettled payments and value of unsettled payments, respectively) and slightly lower for Model 2 (dependent variable value of unsettled payments). The additional explanatory value of the network indicators at node and network level is however very low. With respect to the interaction between network structure and network stability, we conclude that the position of a stricken node within the network has an impact on network stability in the face of an operational shock, although the explanatory value is small. The results for network indicators at network level raise serious doubt about the hypothesis that variations in network structure (within a given payment system) are relevant for network stability.

References

Albert, R – Jeong, H – Barabasi, A-L (1999) **Diameter of the World Wide Web.** Nature 401, 130–131.

Albert, R – Jeong, H, – Barabasi, A-L (2000) **Error and attack tolerance of complex networks.** Nature 406, 378–381.

Albert, R – Barabasi, A-L (2002) **Statistical mechanics of complex networks.** Reviews of Modern Physics 74, 47–97.

Arellano, M (2003) **Panel Data Econometrics.** Cambridge University Press, Cambridge.

Baltagi, B H (2001) **Econometric Analysis of Panel Data.** John Wiley & Sons, Chichester.

Beck, N – Katz, J N (1995) **What to do (and not to do) with time-series cross-section data.** American Political Science Review, Vol. 89, No. 4, 634–647.

Borgatti, S P (2005) **Centrality and network flow.** Social Networks 27, 55–71.

Boss, M – Elsinger, H – Summer, M – Thurner, S (2004) **An empirical analysis of the network structure of the Austrian interbank market.** OeNB Financial Stability Review 7, 77–87.

Breusch, T – Pagan, A (1980) **The LM test and its Applications to Model Specification in Econometrics.** Review of Economic Studies, Vol. 47, No. 1, 239–253.

DeGroot, M H (1985) **Probability and Statistics.** Second Edition, Addison-Wesley. Reading, Massachusetts.

Frees, E W (1995) **Assessing cross-sectional correlations in panel data.** Journal of Econometrics, Vol. 69, No. 2, 393–414.

Friedman, M (1937) **The use of ranks to avoid the assumption of normality implicit in the analysis of variance.** Journal of the American Statistical Association, Vol. 32, No. 200, 675–701.

Inaoka, H – Ninomiya, T – Taniguchi, K – Shimizu, T – Takayasu, H (2004) **Fractal Network derived from banking transaction – An analysis of network structures formed by financial institutions.** Bank of Japan Working papers No. 04-E-04.

Kyriakopulos, F – Thurner, S – Puhr, C – Schmitz, S W (2009) **Network and eigenvalue analysis of financial transaction networks.** European Physical Journal B (forthcoming).

Latzer, M – Schmitz, S W (eds.) (2002) **Carl Menger and the Evolution of Payment Systems.** From Barter to Electronic Money, Edward Elgar, Cheltenham.

Leinonen, H (ed.) (2005) **Liquidity, risk and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies E:31, Helsinki.

Newman, M E J (2003) **The structure and function of complex networks.** (available at http://arxiv.org/abs/cond-mat/0303516).

Newman, M E J (2005) **Power Laws, Pareto Distributions, and Zipf's Law.** Contemporary Physics 46, 323–351.

Oesterreichische Nationalbank and Finanzmarktaufsichtsbehörde (2004) **The Austrian Financial Markets.** Vienna.

Pesaran, M H (2004) **General diagnostic tests for cross-sectional dependence in panels.** University of Cambridge, Cambridge Working Papers in Economics 0435.

Schmitz, S W – Puhr, C (2006) **Liquidity, Risk Concentration and Network Structure in the Austrian Large Value Payment System.** Available at SSRN: http://ssrn.com/abstract=954117.

Schmitz, S W – Puhr, C (2007) **Risk concentration, network structure and the assessment of contagion in the Austrian large value system ARTIS.** In: Leinonen, H (ed.), Simulation studies of the liquidity needs, risks and efficiency in payment network, Bank of Finland Scientific Monograph E:39, Helsinki,183–226.

Schmitz, S W – Puhr, C – Boss, M – Krenn, G – Metz, V (2008) **Systemically Important Accounts. Network Topology and Contagion in ARTIS.** Available at SSRN: http://ssrn.com/abstract=1137864.

Schmitz, S W – Wood, G E (eds.) (2006) **Institutional Change in the Payments System and Monetary Policy.** Routledge, London.

Soramäki, K – Bech, M L – Arnold, J – Glass, R J – Beyeler, W E (2006) **The Topology of Interbank Payment Flows.** Federal Reserve Bank of New York Staff Report, New York No. 243.

Soramäki, K – Beyeler, W E – Bech, M L – Glass, R J (2007) **New approaches for payment system simulation research.** In: Leinonen, H (ed.) Simulation studies of the liquidity needs, risks and efficiency in payment network, Bank of Finland Scientific Monograph E:39, Helsinki, 15–40.

Wooldridge, J M (2002) **Econometric Analysis of Cross Section and Panel Data.** MIT Press, Cambridge.

Zhou, H (2003) **Distance, dissimilarity index, and network community structure.** Physical Review E:67, 061901, 1–8.

Appendix 1

Test results

The assumption of conditional homoskedasticity is tested by a likelihood ratio (LR) test which compares the log-likelihoods for the restricted model (homoskedastic errors) and the unrestricted model (heteroskedastic errors). Both models are estimated by iterated generalised least squares (IGLS). The tests statistics clearly reject the assumption of conditional homoskedasticity for all three models in specification (1). (This is to be expected, as some scenarios hardly generate contagion, so that the variance is extremely low.) The resulting test statistics and error probabilities are shown in table A6.1.

Table A6.1 **Test results – likelihood ratio test for conditional homoskedasticity**

	LR $Chi^2(62)$	Prob.
Model 1	18501	0.00
Model 2	103015	0.00
Model 3	74980	0.00

Source: Own calculations based on data and model specifications as presented in Section 6.5. Likelihood Ratio (LR) and Error Probability (Prob.).

To test for the assumption of serial independence we conduct a Wooldridge test on all three models in specification (1). The test is based on residuals of regressions in first differences, which are then regressed on their lagged values at t-1. The test is robust to conditional heteroskedasticity. The test statistics reject the assumption of serial independence for Models 1 and 3 in specification (1). The resulting test statistics and error probabilities are shown in table A6.2.

Table A6.2 **Test results – wooldrige test**

	F(1,62)	Prob.
Model 1	20.3	0.00
Model 2	2.3	0.14
Model 3	13.5	0.00

Source: Own calculations based on data and model specifications as presented in Section 6.5. Error Probability (Prob.).

Three tests are available to check for cross-panel independence. They were suggested by Friedman (1937), Frees (1995), and Pesaran (2004), respectively. All three test statistics reject the assumption of cross-panel independence for all three models in specification (1) (Table A6.3).

Table A6.3 **Test results – friedman, fress, and pesaran tests**

	Friedman	Prob.	Frees	Prob.	Pesaran	Prob.
Model 1	11728.05	0.00	11.12	0.00	363.11	0.00
Model 2	7378.70	0.00	7.06	0.00	147.08	0.00
Model 3	5744.16	0.00	4.81	0.00	120.80	0.00

Source: Own calculations based on data and model specifications as presented in Section 6.5. Error Probability (Prob.).

The data exhibit high correlation between time-invariant unobservable scenario specific effects ν and the explanatory variables. Consequently, we test for fixed versus random effects. Table A6.4 presents the results of Breusch-Pagan (1980) likelihood ratio (LR) tests of random effects for all three models in specification (1).

Table A6.4 **Test results – Breusch-Pagan likelihood ratio test**

	LR	Prob.
Model 1	306000	0.00
Model 2	43300	0.00
Model 3	230000	0.00

Source: Own calculations based on data and model specifications as presented in Section 6.5. Likelihood Ratio (LR) and Error Probability (Prob.).

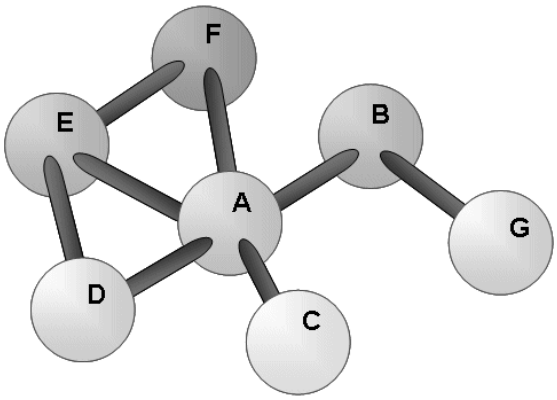
Given the results of all four tests, an ordinary least squares (OLS) estimate of a standard fixed-effects model would yield inconsistent and biased standard errors. We need to apply an estimator that can handle conditional heteroskedasticity, serial correlation, and cross-sectional dependence of the error terms ε.

Appendix 2

Definition of network indicators

This appendix summarises the definitions and formulas for the network indicators used in the paper. In addition, we provide an illustrative example of a network which allows us to visualise different values of network indicators.

Figure A6.1 **A simple network example**



Source: Own calculations.

The network as provided in Figure A6.1 is an undirected, unweighted network with seven nodes (*number of nodes* n=7) and 8 undirected, unweighted links (*number of links* m=16[40]). Table A6.5 summarizes the relevant network indicators from our paper for this network.

[40] There are eight connections in our network, between node i and j, and each of these can be seen as a link from node i to node j as well as from node j to node I, hence m, the total number of links is 16.

Table A6.5 **A simple network example – node-level network indicators**

	A	B	C	D	E	F	G	Network
(Average)* Degree	5	2	1	2	3	2	1	2.3
(Average)* Connectivity	83.3%	33.3%	16.7%	33.3%	50.0%	33.3%	16.7%	38.1%
(Average)* Clustering Coefficient	20.0%	0.0%	0.0%	100.0%	66.7%	100.0%	0.0%	41.0%
Average Path Length	1.2	1.7	2.0	1.8	1.7	1.8	2.5	1.8
(Average)* Betweenness Centrality	85.7%	35.7%	0.0%	0.0%	7.1%	0.0%	0.0%	18.4%
(Average)* Dissimilarity Index	1.5	1.6	1.5	1.4	1.5	1.4	3.1	1.7

*) While the node level indicator of node h is not an average, the corresponding indicator on the network level is indeed the average across all nodes n.
Source: Own calculations.

Measures of Network Structure[41]

The degree k_h of node h is measured by the number of links originating (out-degree) or terminating (in-degree) at node h. In our case of an undirected network, the in- and out-degrees are identical. Take for instance node A, which has links to nodes B, C, D, E, and F, hence node A has a degree $k_A = 5$ while the respective value for node B is $k_B = 2$; ie node A is linked to five other nodes and B to just two. As the network is undirected, degree k remains the same for node A and B independent of whether we look at links that originate or terminate at each node.

On the network level, the average degree k of the network is calculated by summing across all links originating from each node (out-degree k_i^{out}) or terminating at each node (in-degree k_i^{in}) and than averaging across nodes $k = \frac{1}{n}\sum_i k_i^{out} = \frac{1}{n}\sum_i k_i^{in} = \frac{m}{n}$. In our example the average degree of the network is 2.3, based on dividing 16 direct links m by our seven nodes n.

The **connectivity** of node h is its degree over the number of nodes n. In our example, node A has connectivity = 83% while the respective value for node B is connectivity = 33%; ie connectivity puts the absolute value of degree in relation to the size of the network (as measured by the number of nodes n). On a network level, average connectivity is defined by the number of actual (directed) links m over

[41] Where possible, we use the notation of Albert and Barabasi (2002), Soramäki et al (2006) and Zhou (2003).

the number of possible (directed) links n(n–1). In our example the average connectivity of the network is 38.1%, based on dividing 16 direct links m by 42 potential (directed) links between our n nodes.

The **clustering coefficient** $C_C(h)$ of an individual node h with k_h neighbours measures how well the latter are connected with each other. The number of potential links between the k_h neighbours is $k_h(k_h–1)/2$. Let the actual number of nodes between them be E_h, so that $C_C = \frac{E_h}{k_h(k_h - 1)/2}$.

The clustering coefficient of node A $C_C(A) = 20\%$ and that of B $C_C(B) = 0\%$. Ie node A has five direct neighbours, so that the potential number of direct links is ten, but only two direct links exist (D-E and E-F). Therefore, the clustering coefficient of A is 20%. B has two neighbours with one possible direct link, but A and G are not linked, so that the clustering coefficient is 0%.

The average clustering coefficient of network C_C is the average of all individual nodes' clustering coefficients $C_C(i)$ and is hence defined as $C_C = \frac{1}{n}\sum_i C_C(i)$. In our example the average degree is 41.0%, based on our n nodes individual clustering coefficients.

An indicator of the distance d_{ij} between nodes is the lowest possible number of links that connects each node i with each other node j in the network. It is referred to as shortest path length.

We calculated the **average path length** l_h for the originating node h by averaging over the shortest path lengths to each terminating node i. Therefore l_h is defined as $l_h = \frac{1}{n-1}\sum_{h \neq i} d_{hi}$.

In the example the average path length of node A is much lower ($l_A = 1.2$) than that of node B ($l_B = 1.7$), ie from node A any other node can be reached on the shortest path via an average of 1.2 links, while it takes 1.7 links from node B.

For the entire network, the average path length l is defined as the average path length over all originating nodes l_i divided by our seven nodes n, formally written as $l = \frac{1}{n}\sum_i l_i$.

The **betweenness centrality** $C_B(h)$ of node h provides a measure of how many shortest paths d_{ij} pass through node h. Let $s_{ij}(h)$ be the number of shortest paths between all pairs of nodes i and j that pass through node h and let s_{ij} be the number of all shortest paths between all pairs of nodes i and j; then

$$C_B(h) = \sum_{h \neq i \neq j} \frac{s_{ij(h)}}{s_{ij}}$$

In our example there are 44 shortest paths. The lower boundary is given by 42 possible (directed) links n(n–1). To these 42 we must add another 2, as the shortest path from D-F and vice versa could either pass through A or E. By definition, we then have to exclude those 16 paths that link directly with neighbouring nodes, which leaves 28 shortest paths in the denominator. In our example 24 of those pass through node A and ten pass through node B. Therefore their betweenness centralities are given by $C_B(A) = 85.7\%$ and $C_B(B) = 35.7\%$.

For the entire network, the average betweenness centrality C_B is defined as the average of all individual nodes' betweenness centralities $C_B(i)$ and is thus $C_B = \frac{1}{n} C_B(i)$. In our case the average betweenness centrality is 18.4%, based on our n nodes individual betweenness centrality.

Finally, the **dissimilarity index** of two neighbours nodes i and j is defined as

$$\Delta_{ij} = \frac{\sqrt{\sum_{h \neq i,j}^{N} \left[d_{ih} - d_{jh}\right]^2}}{(N-2)}$$

where d_{ih} and d_{jh} are distance measures from nodes i and j to node h. It provides a comparison of the view of the entire network from the perspective of all pairs of neighbouring nodes. For the entire network the average dissimilarity index is $\Delta = \frac{1}{n(n-1)/2} \Delta_{ij}$.

Our example serves to illustrate that, although for most other indicators node A is distinctly different from all other nodes in the network (particularly regarding our other 'advanced' network indicator, betweenness centrality), the view on the rest of the network is very similar, as A is the central node in a cluster. Hence, its dissimilarity index of 1.5 aligns it more or less with all of its direct neighbours. Node F, whose remote position may have gone unnoticed so far (at least based on other network indicators), is shown to be vastly different from the other nodes, however. These distinct features of betweenness centrality and dissimilarity index are also the reason why we introduced them in the first place.

Chapter 7

Simulating the impact of a hybrid design on the efficiency of large-value payment systems

Kemal Ercevik – John Jackson

7 Simulating the impact of a hybrid design on the efficiency of large-value payment systems

Abstract

This paper uses a simulation methodology to quantify the impact of introducing one particular type of 'hybrid design', a centralised receipt-reactive queue, on the liquidity demands faced by banks using large-value payment systems. Using real payments data for the UK we show that significant liquidity savings are achievable if banks choose to submit a high proportion of their payments into the queue. The relationship between values queued and savings achieved is found to be positive and non-linear. Liquidity savings are distributed unevenly; the largest users do not benefit significantly while the smaller users make significant savings. A synthetic payments dataset is used to demonstrate that these results hold more generally. The level of liquidity recycling in the existing payment system is shown to be a key determinant of the impact of hybrid design.

7.1 Introduction

This paper uses real payment data to test the possibility that users of large value payment systems could benefit from the introduction of a hybrid payment system design. Simulations are used to quantify the impact of such a design change on users in terms of the impact on both their liquidity usage and the degree of settlement delay introduced. The first part of the analysis is carried out using historical data from CHAPS Sterling, the UK's large-value payment system. Our analysis is then extended by generating synthetic payments data and using it to probe the extent to which our findings apply to large-value payment systems more generally.

CHAPS Sterling is the UK's large value payment system (LVPS) with a mean daily turnover of £284 billion.[1] It operates on a real-time

[1] Payment Systems Oversight Report 2008 (www.bankofengland.co.uk/publications/psor/psor2008.pdf).

gross settlement (RTGS) basis: payments are processed transaction by transaction with simultaneous debit of the payer and credit of the receiver in accounts held at the Bank of England. RTGS eliminates interbank credit risk by providing immediate finality of payments: once a transaction has settled it is irrevocable and cannot legally be unwound (for a detailed explanation of RTGS systems, see BIS, 1997). This important feature of RTGS has led to its adoption in LVPSs worldwide, as banks and central banks have found the settlement risk inherent in deferred net settlement (DNS) systems to be too high given the size of interbank exposures that can occur. Although RTGS in central bank money eliminates settlement risk, it can have the undesired consequence of increasing the cost of making payments and potentially increasing liquidity risk faced by banks. When payments are settled gross rather than being netted out at the end of the day, banks typically require more liquidity to make their payments. CHAPS banks, for example, require on average three times more liquidity under RTGS than they would have needed under a DNS system with multilateral end-of-day netting.[2]

The past decade has seen a growing trend of adoption of hybrid payment system designs for the settlement of large-value payments in developed countries.[3] Hybrid systems seek to be liquidity-efficient without introducing significant amounts of settlement risk by combining features of both RTGS and DNS systems. A defining feature of these systems is that payments can be centrally queued, with their release conditional on certain criteria, such as the arrival of offsetting payments. Offsetting occurs when two or more payments are settled simultaneously. Although in legal terms settlement is still gross (ie each transaction is settled with finality individually) payments that are offset can be thought of as having generated their own liquidity, as offsetting has the same economic effect as the netting of payments.

Two main types of hybrid systems have emerged to date. One type, which has evolved from DNS, is called continuous net settlement (CNS), the CHIPS system in the US is the leading example of a CNS design. Although settlement risk is significantly reduced in CNS systems (payments that are offset and settled with finality in batches intraday are not dependent on the subsequent settlement of

---

[2] Source: Bank of England payments database and authors' calculation for December 2006.

[3] Calculations in Bech et al (2008) show that in 1999 only 3% of wholesale payments by value in CPSS countries were settled over payment systems with a hybrid design, while in 2005 that had risen to 32%.

other payments in the system), it is not completely eliminated. The second type of hybrid system incorporates a queuing facility into RTGS, in effect creating multiple streams into which banks can channel their payments: typically a time-critical (RTGS) stream and an offsetting (queuing) stream. Examples of systems with this type of functionality are TARGET2, the centralised LVPS of the Eurosystem, and BOJ-NET. Such queue-augmented RTGS systems do not give rise to settlement risk,[4] as all queued payments are legally considered to be final at the time they are offset. For a more detailed explanation of hybrid system designs see McAndrews and Trundle (2001) and BIS (2005).

Although queue-augmented RTGS systems can help reduce liquidity risk without introducing settlement risk, they can introduce settlement delays which may impose additional costs on banks. A number of theoretical and empirical papers assess this fundamental trade-off between liquidity efficiency and settlement delay in an attempt to find the optimal settlement arrangement for LVPSs. Johnson, McAndrews and Soramaki (2004) use simulation techniques to assess the impact of various types of hybrid functionality on Fedwire, the RTGS system used to settle large value US dollar payments. The paper finds that one mechanism in particular, receipt-reactive gross settlement (RRGS), can potentially reduce participants' costs of obtaining intraday credit, whilst only modestly delaying the average time of settlement.[5] RRGS is a novel queue release mechanism proposed by the paper that conditions the settlement of queued payments on the arrival of incoming payments. This feature ensures that all the liquidity posted by a bank is reserved solely for making time-critical payments. Their paper recognises that the introduction of RRGS functionality would provide a good incentive for banks to submit payments earlier in the day, but does not attempt to incorporate this behaviour in its simulations.

Willison (2004) and Martin and McAndrews (2008) both use theoretic models to predict and compare equilibria for RTGS and hybrid system designs. Willison examines the trade-off between cost of liquidity and operational risk caused by payment delay, and finds

[4] As long as receiving banks do not anticipate payments they are due to receive in the queue and credit beneficiaries' accounts before the payment has been settled / offset with finality. To avoid this, some hybrid systems provide very limited information about payments in the central queue (payer, payee, value only) until finality has been achieved.

[5] This cost reduction is based on the Federal Reserve's method of charging for intraday credit, which is a fee based on banks' average overdrafts calculated at the end of each calendar minute, and does not necessarily apply to a system where intraday credit is free but collateralised, such as CHAPS.

that the first-best solution is unattainable under RTGS. He also finds that a hybrid payment system design outperforms RTGS when payments can be offset either in the morning or all day. Martin and McAndrews compare welfare among RTGS, receipt-reactive and balance-reactive hybrid system designs (a balance-reactive system is one where settlement of queued payments is conditional on a participant's account balance). They find that a balance-reactive system can provide higher or lower welfare than RTGS depending on certain criteria, such as the cost of delaying payments in the system or the probability of participants facing liquidity shocks (modelled as an unexpected fall in their account balances as a result of a time-critical payment to an ancillary settlement system). Welfare is defined in terms of the cost of liquidity and cost of delay borne by the participants. A receipt-reactive system on the other hand is found to weakly dominate RTGS: it can achieve a level of welfare at least as high as, if not better than, RTGS. Finally, a receipt-reactive system can provide higher or lower welfare than a balance-reactive one depending on the cost of delaying payments in the system and the probability of liquidity shocks. They also suggest that by conditioning payments on the receipt of offsetting payments and having some insurance against the risk of having to borrow costly intra-day funds from the central bank, participants would be more willing to submit payments earlier under a hybrid system design than they would under RTGS.

The purpose of this paper is to investigate the possibility of achieving liquidity savings in CHAPS, and other large-value payment systems, by simulating the impact of receipt-reactive hybrid functionality[6] using both historical and synthetic payments data. We attempt to fill some of the existing gaps in the literature outlined above by:

– Analysing the aggregate impact of hybrid functionality on CHAPS banks in terms of costs of liquidity and degree of delay introduced

[6] CHAPS is not considered to be a hybrid system, although it does feature a central queue. A 'circles' process is run once a day, and can be run additionally by the RTGS system controller at any time during the day in order to resolve gridlock situations when participants have insufficient funds on their accounts, which momentarily locks payments waiting in the central queue and multilaterally attempts to offset as many of them as possible. This is not, however, used as a liquidity saving feature in its current form; banks prefer to queue payments in their internal schedulers rather than submit them to the central queue. A gridlock has never occurred to date due to the posting of ample amounts of liquidity by CHAPS banks at the start of each day.

by RRGS,[7] experimenting with different criteria for time-criticality. We find that significant liquidity savings are achievable where banks choose to submit a very high proportion of payments (> 90%) into the receipt-reactive queue, and show that the relationship between values queued and savings achieved is positive and non-linear. The level of settlement delay introduced also increases in a non-linear fashion as queued values rise, but does not reach excessive levels under any assumptions used.

– Assessing the impact of hybrid functionality at the individual bank level. We show that liquidity savings are unevenly distributed across banks, with an inverse relationship between the value of payments made by a bank and the liquidity saving benefits to that bank from the introduction of RRGS. For the largest CHAPS banks liquidity demands can even rise slightly under RRGS.

– Endogenising banks' likely payment submission behaviour under RRGS by submitting some non time-critical payments earlier in the day. We show that where significant levels of payment delay exist in an RTGS system, the value-weighted average time of settlement can be brought forward significantly by the introduction of receipt-reactive functionality due to the impact on payment behaviour.

– Generating synthetic payment data to probe the extent to which our findings are CHAPS specific and investigate the key determinants of the impact of hybrid functionality. We find that the impact of hybrid functionality is strongly influenced by the structure of payment flows, especially the level of liquidity recycling being achieved under RTGS, the number of direct participants in the system and the amount of payments they process. We corroborate that key aggregate and bank level findings for CHAPS hold more generally. The differences between real and synthetic data in the impact of RRGS on volatility of liquidity demands suggest that CHAPS banks already submit payments using strategies that reduce the volatility of their liquidity needs.

[7] We would also have liked to simulate a balance-reactive hybrid system design. Unfortunately, such a queuing algorithm is not available in the version of the Bank of Finland simulator (BoF-PSS2 v2.2.5) used in this work. Simulating this functionality would be an obvious extension to our work, once the functionality becomes available.

The rest of this paper is organised as follows. In the next section, we explain our simulation methodology by defining RRGS and introducing the metrics we use to measure the impact of hybrid functionality on liquidity usage and payment delay. In Section 7.3, we report results of our simulations, interpreting them in Section 7.4 and examining possible policy implications. We conclude in Section 7.5 by summarising our key results and outlining some possible extensions to our work. More detailed information on how we generate our synthetic payments datasets can be found in the Appendix.

7.2 Method

We run simulations to test the possibility of making liquidity savings in CHAPS by complementing the existing RTGS stream with RRGS, using a month of historical payments data.[8,9] We use the Bank of Finland payment and settlement simulator (BoF-PSS2), which is described in detail in Leinonen and Soramaki (2003). Section 7.2.1 describes the RRGS functionality in more detail. Time-critical payments and the metrics used to measure the impact of our simulations are defined in sections 7.2.2 and 7.2.3 respectively. Sections 7.2.4 and 7.2.5 explain how we endogenise banks' payment submission behaviour and generate synthetic payments data.

[8] December 2006 – a 'clean' month with no CHAPS settlement extensions. 19 working days in total.

[9] In addition to the RRGS algorithm, the Bank of Finland offers a range of bilateral and multilateral offsetting algorithms whose impact on CHAPS could in principal be simulated. In practice, however, this is not a viable alternative as it is not possible to simulate the impact of such an algorithm while continuing to allow time-critical payments to be settled immediately without carrying out two distinct simulations: a pure RTGS simulation for time-critical payments, and a continuous net settlement (CNS) simulation with multilateral offsetting for the remaining transactions. Simulating such an arrangement using CHAPS data results in banks requiring more liquidity than where all the payments are settled RTGS (ie without offsetting) under one simulation. This occurs because liquidity recycling is disrupted by the 'splitting' of transactions between two accounts. A surplus of liquidity in one account cannot be used to fund a deficit in the other, requiring further liquidity injection, which would not have been the case if the funds had been posted onto a single account (as is the case, for example, with the RRGS algorithm which has multiple streams but uses a single account).

7.2.1 Receipt-reactive gross settlement

The RRGS algorithm can be viewed as an extreme case of liquidity reservation functionality. All the liquidity posted by a bank into the system is reserved to allow payments which that bank designates as high priority to be settled immediately. The same bank's low priority payment messages are released for settlement, on a first-in first-out (FIFO) basis, only where they can be settled using liquidity from the arrival of incoming funds within a pre-specified period of time. Johnson et al (2004) use calendar minutes as the time intervals in their paper: in any minute the algorithm allows the release of as many payments from the front of the queue as is possible to offset, but not exceed, the amount of incoming funds received in that minute. They use this approach because in Fedwire charges for banks' daylight overdrafts are calculated at the end of each calendar minute.

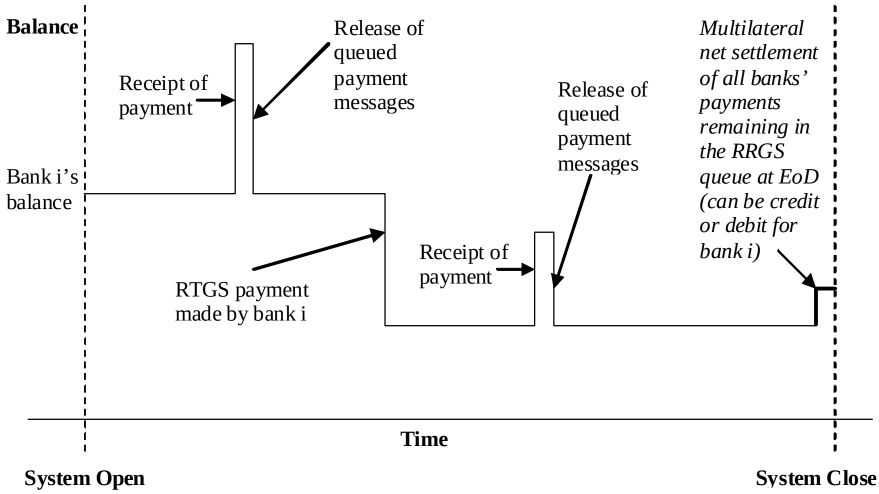
In our simulations it is appropriate to take the entire CHAPS day as one continuous period,[10] as the key determinant of the cost of posting liquidity in CHAPS is the maximum liquidity needed throughout the day. This means that the RRGS algorithm runs on a continuous basis: a payment received at 9am can cause the release of a payment entered into the queue at 10am or even 4pm as long as aggregate payments received by bank i ≥ aggregate queued payments sent by bank i at that point in time, including the queued payment(s) being released. Under RRGS, a bank does not necessarily have to post liquidity to cover the gross value of all its outgoing time-critical payments: incoming funds can be used to finance both time-critical and non-time critical payments.

Finally, it is worth noting that although we refer to the term offsetting throughout the paper, this misleadingly suggests that offset payments are always released for settlement at the same instant. This can and often does happen, but there are also instances where a payment entered into the queue is conditionally released in response to a payment received much earlier in the day. RRGS should therefore be viewed as a mechanism for co-ordinating the immediate settlement of time-critical payments, together with the conditional release of the

[10] We also investigate whether adopting a greater number of distinct periods over which incoming payments are cumulated impacts our results by experimenting with 1 hour and 5 minute intervals. We find that estimated collateral posting by banks is not significantly affected, as mean liquidity requirements across the month slightly decrease, but this is offset by an increase in volatility (standard deviation) across the 19 days: for a detailed explanation of the metrics used to quantify the liquidity burden faced by banks and their impact on estimated collateral postings see Sections 7.2.3.1 and 7.4.4 respectively.

remaining payments against incoming payments: where we use the term offsetting in the context of RRGS, we specifically mean offsetting through conditional release. Also, if there are any unsettled payments remaining in the queue at the end of the day, we assume these are settled on a multilateral net basis. The Figure below illustrates how the RRGS mechanism works.

Figure 7.1 **Dynamics of a bank's balance under RRGS**



Source: Bank of Finland PSS2 User Manual v 2.2.0

7.2.2 Time-criticality

Throughout our analysis we always require time-critical payments[11] to be settled immediately but allow the remaining payments to be queued, waiting for incoming payments to trigger their release. Anecdotal evidence suggests the main payment types of payments which are considered by CHAPS banks to be truly time-critical are

[11] The term time-critical payment refers to a payment for which the sender deems there are significant private or social costs to delay; indeed, in some cases, failure to pay at, or by, a given time intraday may constitute technical default.

payments that are either: (a) to or from the account of CLS;[12,13] (b) extremely large in value; or (c) require prompt settlement by customers eg house purchases. It is not possible to identify the last category of time-critical payments in our dataset but we always treat payments to and from CLS Bank's CHAPS account as high priority and settle them RTGS. As we have no other obvious way of identifying precisely other categories of time-critical payments in our data, we experiment with two approaches to proxy time-criticality:

(i) Payments of a size greater than or equal to a certain value threshold are time-critical and settled immediately; others only have to be settled sometime during that day (at the latest by the end of day) and could therefore be queued and offset. We experiment with different value thresholds by treating payments larger than £100m, £500m and £1bn as time-critical.

(ii) A proportion (r) of payments in our sample are designated as time-critical and settled immediately. This approach is used in Johnson et al (2004). For ease of comparison we adopt values of r to ensure that the same proportion of payments by value is treated as time-critical under approaches (i) and (ii).

In practice, banks' time-critical payments are likely to be somewhere in between the two sets of payments we identify under the approaches above. Our method therefore enables us to report a range of possible liquidity saving / settlement delay trade-offs, from which both extremes of the impact of RRGS on CHAPS banks can be observed.

[12] CLS is the Continuous Linked Settlement system which provides payment versus payment settlement of FX transactions, see www.cls-services.com for more details. Users' sterling pay-ins to CLS are made across CHAPS.

[13] More generally, pay-ins to all ancillary systems are typically viewed as highly time-critical. However, CLS is the only payment of this type which is made across CHAPS. Settlement of BACS, the UK retail clearing system, takes place as a non-CHAPS transfer across banks' RTGS accounts and is not included in our dataset. Similarly, liquidity transfers to CREST for settlement of DvP transactions occur across other liquidity transfer accounts.

7.2.3 Metrics for comparison

Throughout the paper we compare metrics from our simulations with those obtained from a pure RTGS simulation. We focus on the following measurements to quantify the impact of RRGS on the trade-off banks face between liquidity usage and settlement delay.

7.2.3.1 Metrics used to quantify the liquidity burden faced by banks

We use various measures to identify relevant indicators of banks' liquidity posting requirements. One simple measure is the mean of the maximum daily net debit position faced by each bank across the month, which is then summed across banks to get an aggregate figure for the system. In practice it would not be feasible for banks to post this exact amount into the system ex ante as it is determined by the submission behaviour of all banks in the system, but it gives a good indicator of liquidity demands faced by banks. One potential drawback of this metric is that it cannot quantify changes in the volatility of liquidity demands that may be caused by RRGS. To capture this we quantify how the standard deviation of maximum daily net debit positions over the month changes for each bank. We also display results for the liquidity required by each bank to cover the single maximum net debit position faced across the entire month.

Mean of maximum liquidity requirement across the month for bank i

$$\overline{\mathrm{MLR}_i} = \frac{1}{n}\sum_{j=1}^{n}\mathrm{MLR}_{ij}$$

where
n = number of CHAPS days being simulated
MLR_{ij} = maximum liquidity requirement of bank i on day j

Standard deviation of maximum liquidity requirement across the month for bank i

$$\mathrm{Sd}_i = \sqrt{\sum_{j=1}^{n}(\mathrm{MLR}_{ij} - \overline{\mathrm{MLR}_i})^2 /(n-1)}$$

Maximum of maximum liquidity requirement across the month for bank i

$$\mathrm{Max}_i = \max_j(\mathrm{MLR}_{ij})$$

7.2.3.2 Metric used to quantify average settlement time

The basic indicator that we consider to quantify the average settlement time under different simulations is the value-weighted average settlement time (AST), calculated as follows

$$\mathrm{AST} = \frac{\sum_i t_{s,i} a_i}{\sum_i a_i}$$

where
$t_{s,i}$ = settlement time of payment i
a_i = value of payment i

Early settlement is desirable from an operational risk perspective: the earlier payments are settled, the lower the risk of having large amounts of payments remaining to be settled following an unexpected operational outage.

7.2.3.3 Recycling ratio

We use the recycling ratio, r, to calculate the liquidity efficiency of the system before and after the introduction of RRGS. This is based on the method used in Becher et al (2007) and is calculated as follows

$$r = \frac{\sum_{j=1}^{n} \sum_i a_{ij}}{\sum_{j=1}^{n} \sum_i \mathrm{MLR}_{ij}}$$

The recycling ratio measures on average how many times the system recycles a unit of liquidity by comparing the total value of payments submitted with the value of liquidity that needs to be posted by members of the system to allow settlement to occur.

7.2.4 Endogenising payment submission behaviour

Anecdotal evidence obtained from dialogue with CHAPS users explains how banks might alter their behaviour in response to a liquidity saving mechanism such as RRGS. Banks would have an incentive to submit payments to the central scheduler as early as possible to maximise the benefits of offsetting, safe in the knowledge that their liquidity posted at the start of day would be reserved solely for making time-critical payments. This incentive to submit payments earlier in the day is also predicted by recent literature (in Johnson et al, 2004, and implicitly in Willison, 2004).

We test the impact of this prediction by artificially changing the submission times of payments in some of our simulations. However, banks have made it clear to us that their ability to submit payments earlier in the day would depend on having prior knowledge of individual payments, which varies from bank to bank according to the mix of payments they process. We simulate the impact of these behavioural changes by submitting randomly selected 20, 50 and 80% of non-time critical payments greater than £1m at the start of day instead of their original submission times.[14] In doing this, we incorporate overnight loans data obtained via the Furfine method by bringing forward all loan repayments to the beginning of the day (as these are known in advance) and none of the new loans being generated on that day (as these are not known until later in the day).[15]

7.2.5 Synthetic payments

We generate datasets of synthetic payments to test the extent to which our findings are CHAPS-specific (for detailed methodology, see Appendix 1). Using such data allows us to vary the numbers of banks in a system, the numbers of payments being settled, and the distribution of values and volumes of payments across banks to see what drives the impact of RRGS.

[14] Banks have told us that they tend to make their small value payments (we assume this means for value < £1m) immediately as they are received without queuing them in their internal schedulers, reflecting the fact these payments do not impose significant liquidity demands. It would not therefore be realistic to alter the submission times for these payments as they are not subject to delay. Similarly, we do not change the submission times for time-critical payments, since by definition these could not have been delayed.

[15] We identify overnight loan payments in CHAPS using a Matlab program which matches all the overnight loans in our dataset to loan repayments the following day. The program uses a similar methodology to that outlined in Millard and Polenghi (2004).

7.3 Results

We display system-wide results using CHAPS data in Section 7.3.1 (aggregate individual banks excluding Bank of England and CLS Bank[16]) before looking at the impact of RRGS on individual CHAPS banks in Section 7.3.2. We then compare our findings with results obtained using a dataset of synthetic payments in Section 7.3.3.

7.3.1 Aggregate findings

Table 7.1 overleaf shows that no significant liquidity savings or settlement delays are observed where only half of all payments by value are submitted to the receipt-reactive stream. Significant mean liquidity savings start to occur as the proportion of queued payments is raised; at the same time minor increases in settlement delay can be observed. The impact of both effects is increasing in the numbers of payments queued.

Mean liquidity savings are not influenced by the method through which time-critical payments are selected. By contrast, the volatility (standard deviation) of liquidity demands is affected. Volatility is unchanged or even increases slightly under RRGS where a value threshold is used, while volatility falls where a volume-based selection method is adopted. This difference in volatility is also evidenced by a corresponding difference in liquidity savings based on the maximum liquidity requirement measure.

Earlier submission to a central queue of a subset[17] of non-time critical payments has little impact on observed liquidity savings under RRGS, but can significantly bring forward the average time of settlement when compared to observed settlement times in CHAPS.

[16] We treat RBS and NatWest as a single entity throughout our simulations, even though they have separate CHAPS Sterling settlement accounts. We therefore report findings for only 12 banks in our results even though there are 15 direct CHAPS Sterling participants in our dataset.

[17] Based on certain criteria – see Section 7.2.4 for a detailed explanation.

Table 7.1 **Impact of RRGS on CHAPS at the aggregate level**

Time-critical payments[a]		Early submission[c]	% Δ Liquidity requirement			Δ Settlement time	
Criteria	Proportion[b]		Mean	St dev	Max	AST[d] hh:mm	St dev[e] hh:mm
≥£100mn	54%	-	-2	+1	0	+00:01	00:00
≥£500mn	12%	-	-10	+10	+3	+00:12	-00:01
≥£1bn	4%	-	-38	0	-9	+00:37	+00:09
Random 50%	51%	-	-1	-2	-1	+00:01	00:00
Random 10%	11%	-	-12	-10	-8	+00:11	-00:01
Random 3%	4%	-	-37	-18	-16	+00:25	+00:04
≥£100mn	54%	Random 50%	0	+5	+1	-00:57	+00:01
≥£500mn	12%	Random 50%	-15	+9	0	-01:20	+00:06
≥£1bn	4%	Random 50%	-37	+5	-4	-00:33	+00:30
Random 10%	11%	Random 50%	-14	-12	-5	-01:15	+00:02
Random 3%	4%	Random 50%	-39	-18	-16	-01:01	+00:12
≥£1bn	4%	Random 20%	-37	+2	-8	+00:05	+00:15
≥£1bn	4%	Random 80%	-37	+2	-7	-01:01	+00:43

a CLS payments always treated as time-critical.
b Proportion of all payments that are time-critical (by value).
c A dash indicates that original (RTGS) payment submission times have not been altered. In subsequent simulations we extend our analysis by endogenising banks' payment submission behaviour: banks submit all overnight loan repayments and a randomly selected % of payments (which are ≥ £1mn and non-time critical in both cases) to the RRGS queue at the start of day. Submission times of new overnight loans are left unchanged.
d Change in value-weighted Average Settlement Time (AST) compared to RTGS AST of 11:37am.
e Change in standard deviation of settlement time across the month compared to RTGS.

7.3.2 Bank level findings

Table 7.2 overleaf shows disaggregated results for value (≥£1bn) and volume (random 3%) based time-criticality thresholds, grouping banks by the mean liquidity savings they experience when using RRGS. We see that the mean liquidity requirements of the largest two banks increase slightly as a result of the introduction of RRGS. Between them these banks settle half the value of both total payments and of time-critical payments. They both have high liquidity recycling ratios under RTGS.

We see another small group of medium sized banks settling an average of 10% of value and of time-critical payments, which experience moderate mean liquidity savings. Their liquidity recycling ratios prior to RRGS are also moderately high.

Finally, the majority of banks are much smaller in terms of total value and proportion of time-critical payments settled (around 3–4%). They typically have very low recycling ratios under RTGS, and benefit the most from the introduction of an RRGS stream. This pattern of mean liquidity savings being broadly inversely correlated with recycling ratios under RTGS and the sizes of banks' payment flows, is observed for both value and volume based time-criticality thresholds.

Table 7.2 **Impact of RRGS on CHAPS at grouped bank level**

Δ Mean liquidity requirement	# Of banks	Avg value settled	Avg proportion of time-critical payments settled	Recycling ratio	
				RTGS	RRGS
≥£1bn time-critical					
$MLR_i > 0\%$	2	27%	25%	30	28
$0\% > MLR_i > -40\%$	2	9%	13%	14	21
$MLR_i < -40\%$	8	3%	3%	9	24
Random 3% time-critical					
$MLR_i > 0\%$	2	27%	30%	30	26
$0\% > MLR_i > -40\%$	1	10%	8%	17	25
$MLR_i < -40\%$	9	4%	4%	9	23

7.3.3 Results using a synthetic payments dataset

Similar simulations were run using our synthetic dataset to attempt to replicate the results found using CHAPS data (ie those displayed in the first six rows of Table 7.1 and in Table 7.2). Baseline results are produced for simulations with 10 banks, with the value profile of payment flows and the distribution of payment volume across banks both being drawn from log normal distributions, and with a process of squaring-off of balances taking place during the second half of the payment day. Sensitivity analysis is carried out to examine how the impact of RRGS is altered where the number of banks making payments is increased (we simulate systems with 100 and 1000 banks).

Our results can be found in Table 7.3. We observe some clear similarities between these results and those displayed in Table 7.1. In particular, no significant liquidity savings are observed where only half of all payments are put into the receipt-reactive stream, but significant mean savings do occur as the proportion of queued payments is raised, with the effect increasing in the value of payments queued. Mean savings are of the same magnitude under value and volume based thresholds for time-critical payments.

One key difference is observed. With the synthetic payments dataset, the volatility of liquidity demands decreases by a greater amount than the decrease in mean liquidity demands for both time-criticality thresholds. By contrast, volatility falls by less than the mean (and in the case of value based thresholds doesn't fall at all) with our CHAPS dataset.

Our sensitivity analysis shows that there is a strong correlation between the number of banks in the payment system and liquidity savings seen under RRGS, with the largest savings observed in our dataset with 1000 banks. This seems to be linked to the observation that recycling ratios fall as the numbers of banks in the system is increased.[18]

We also try a 10 bank simulation where we reduce the overall number of payments from 20,000 per day to 5,000. This halves the system's recycling ratio under RTGS (from 47 to 23), possibly due to the probability of recycling payments falling when there are fewer payments to be made in the system. Correspondingly, the benefit of RRGS doubles on the mean liquidity requirement measure (from 17% to 37%), and the recycling ratio of the system improves by 50% after RRGS (from 23 to 36).

[18] Simulating a uniform distribution of bank sizes and payment values has a similar effect to increasing the number of banks: the less concentrated payments are among a few participants, the more opportunity there is for liquidity savings to be made from co-ordinated settlement of non-urgent payments.

Table 7.3 **Impact of RRGS using artificial payments datasets**
Log-normal distribution of payment values and bank sizes, 20 days, 400,000 payments

Time-critical payments	Number of banks	Recycling ratio		% Δ Liquidity requirement		
		RTGS	RRGS	Mean	St dev	Max
Largest 50%	10	47	47	0	0	0
Largest 10%	10	47	50	-7	-7	-7
Largest 5%	10	47	56	-17	-14	-17
Random 50%	10	47	47	-1	-2	-1
Random 10%	10	47	50	-8	-14	-14
Random 5%	10	47	58	-19	-28	-22
Largest 5%	100	20	27	-28	-33	-29
Random 5%	100	20	28	-30	-50	-38
Largest 5%	1000	11	20	-45	-44	-40
Random 5%	1000	11	21	-46	-68	-59
100,000 payments						
Largest 5%	10	23	36	-37	-30	-33

Figure 7.2 below shows how the value of payments submitted is distributed across the dataset of 1000 banks. Figure 7.3 illustrates how individual recycling ratios are linked to a bank's size, as defined by the aggregate value of payments sent. Both Figures display the y-axis on a log scale.

Figure 7.2 **Average value submitted (1000 banks)**
Bank sizes and payment values log-normally distributed

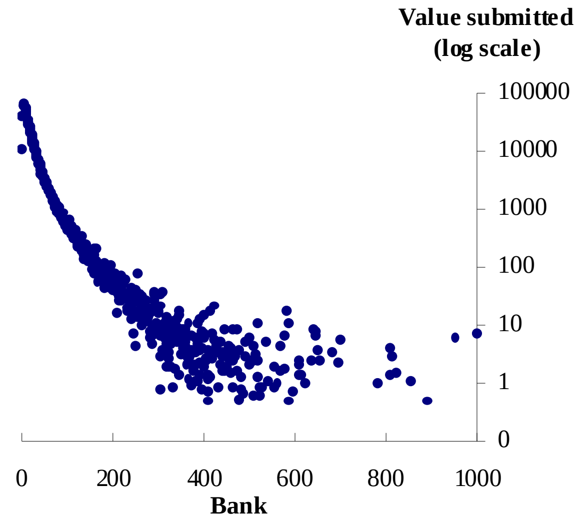


Figure 7.3 **RTGS recycling ratios (1000 banks)**
Bank sizes and payment values log-normally distributed

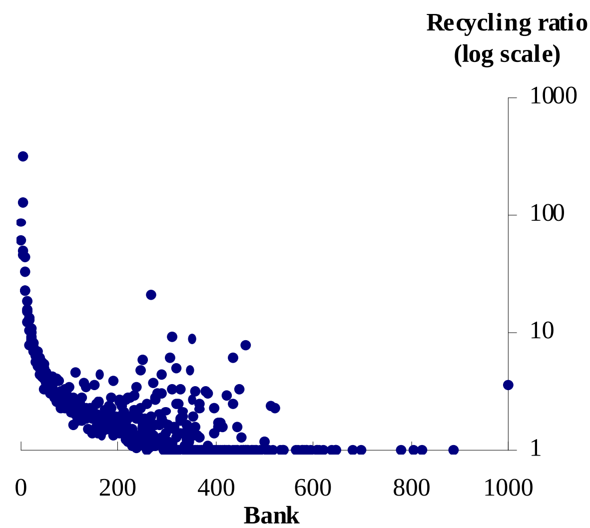
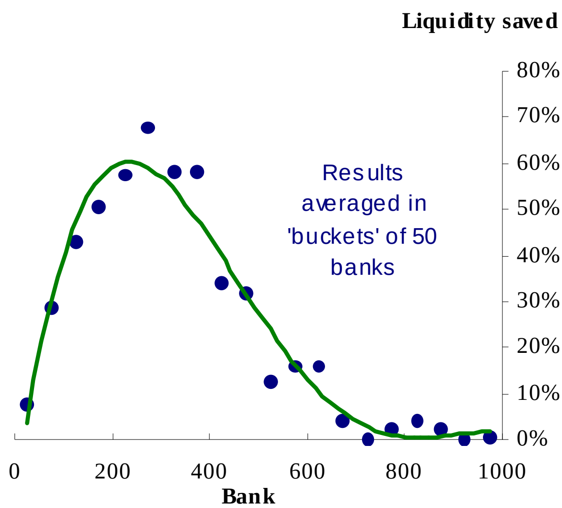


Figure 7.4 shows mean liquidity savings under RRGS where the largest 5% of payments are treated as time-critical, by averaging banks in 'buckets' of 50. Although not directly observable in the Figure due to averaging, those banks with the highest recycling ratios under RTGS see their liquidity demands increase under RRGS, the same pattern observed in our CHAPS data in Table 7.2. Again we see smaller banks with lower recycling ratios under RTGS making larger savings. The one caveat to this pattern is that the very smallest banks in the sample often do not see any savings as their volumes of payments are not sufficiently large for RRGS to give any benefit (ie on a typical day they do not have any offsetting benefits to take advantage of).

Figure 7.4 **Mean liquidity savings under RRGS**
Bank sizes and payment values log-normally distributed – largest 5% of payments time-critical



7.4 Interpretation

This section discusses some of the key findings reported in the previous section. Section 7.4.1 discusses aggregate findings from CHAPS data. The non-linear increase in aggregate liquidity savings seen as more payments are queued is examined in more detail. In addition the differing impact of volume and value based time-criticality thresholds are discussed. Section 7.4.2 examines why significant differences in liquidity savings are observed across CHAPS banks. In section 7.4.3 the bias that might be introduced because our simulations cannot capture the use of bilateral monitoring by CHAPS banks is discussed, with particular reference to its impact on the average time of settlement measure. Section 7.4.4 considers how liquidity savings observed using CHAPS data might be translated into cost savings, using regression analysis borrowed from James and Willison (2004). Section 7.4.5 identifies general policy implications of our findings, drawing on similarities and differences observed in the results obtained between the real and generated payments datasets.

7.4.1 Aggregate findings

Figure 7.5 **Non-linear profile of mean liquidity savings: value based threshold**

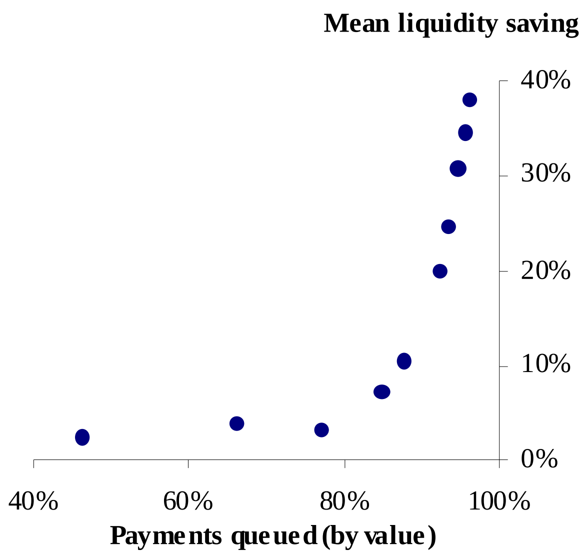
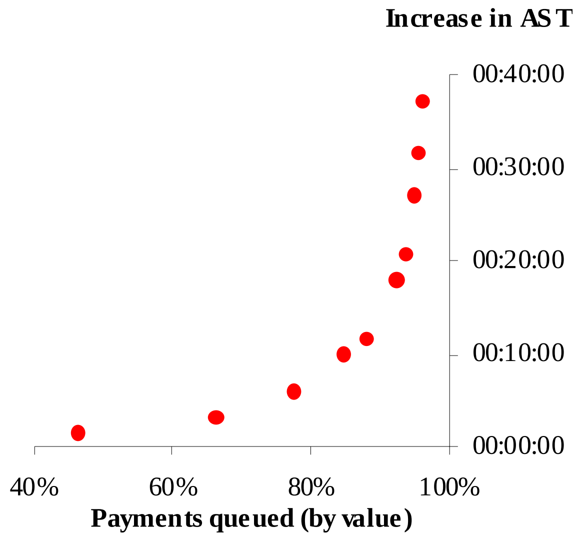


Figure 7.6 **Non-linear profile of increase in average settlement time (AST): value based threshold**



Figures 7.5 and 7.6 above illustrate the non-linear relationship between values queued and the liquidity saving and average settlement time metrics. The nature of this relationship suggests that the benefits of RRGS depend on having a critical mass of payments whose submission is being coordinated by the queue-release algorithm; for CHAPS this point is reached where 80–90% of payments by value are queued. It seems likely that the proportion of queued payments beyond which RRGS starts to offer significant savings will vary depending on the characteristics of the payment system under investigation, for example we might expect that where a system's aggregate recycling level under RTGS is significantly lower than that seen in CHAPS, the effect would emerge at lower proportions of payments. Testing this proposition would be a useful extension to our analysis.

Although our aggregate results show that mean liquidity savings are similar under the value and volume based time-criticality thresholds (Section 7.3.1 – Table 7.1), we observe significant differences in the volatility of these savings across the two methods. The standard deviation increases when using the value based approach but decreases (although by less than the mean) under the volume based approach. The Figures below may provide some insight into this difference. Figure 7.7 shows the distribution of time-critical payments through the day by aggregating all payments sent in half an hour intervals, and taking the mean across the month. As shown by the blue and red bars, the mean profiles for the value and volume based thresholds are very similar, which is consistent with our finding that mean liquidity requirements are not influenced by the method of selecting time-critical payments. In contrast, we see that the standard deviation of time-critical payments across the month is higher (and more volatile through the day) under the value based threshold. One likely factor here is that the number of payments classified as time-critical is much lower under the value based threshold than the volume based threshold. It seems plausible that this translates into a greater volatility in the liquidity demands under RRGS because it makes the timing and destination of payments which provide the liquidity to initiate the settlement of queued payments more unpredictable.

Figure 7.7 **Value profile of time-critical payments in half hour intervals: mean across 19 days**

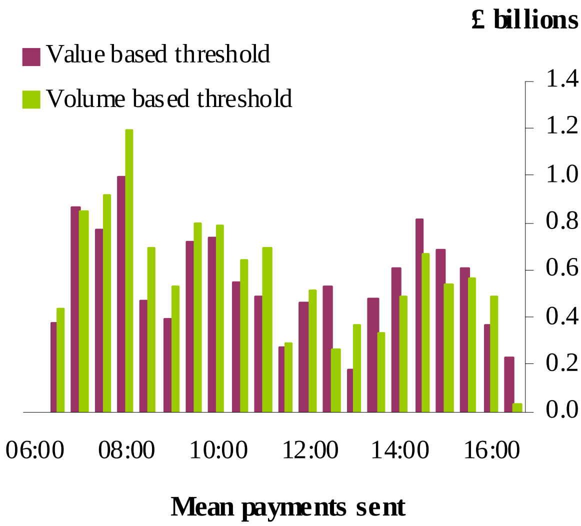
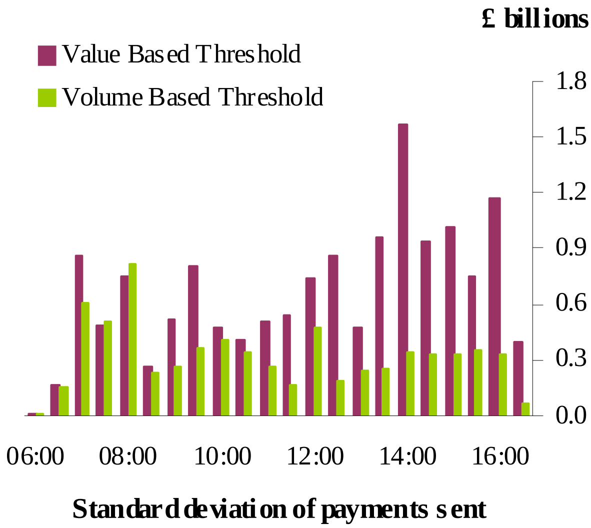


Figure 7.8 **Value profile of time-critical payments in half hour intervals: s.d. across 19 days**



7.4.2 Distribution of RRGS benefits on individual banks

Table 7.2 in section 7.3.2 shows large variations in the impact of RRGS on individual banks. In this section, we explain the reasons for these variations.

Why does liquidity requirement under RRGS increase for some banks?

Figures 7.9–7.12 overleaf display the profile of payments sent (netted against incoming payments) before and after RRGS for two banks: A and B. Bank A's mean liquidity requirement slightly increases under RRGS, whereas Bank B is one of the biggest beneficiaries, seeing a large decrease in its mean liquidity requirement.

It appears that RRGS can potentially disrupt the recycling of payments for banks who already use liquidity very efficiently under RTGS, such as Bank A. Payments which they were due to receive from other banks, and subsequently use to fund their outgoing payments, are now being queued in the central scheduler (see the circled sections of Figures 7.9 and 7.10 below). Since they must make their time-critical payments without delay, they end up using more of their own liquidity to fund their outgoing payments under RRGS. In contrast, banks whose payment profiles are not as liquidity efficient under RTGS, such as Bank B, benefit the most from RRGS (Figures 7.11 and 7.12).

Figure 7.9 **Profile of payments sent under RTGS (net of incoming payments) for Bank A: mean value at half an hour intervals across the month**

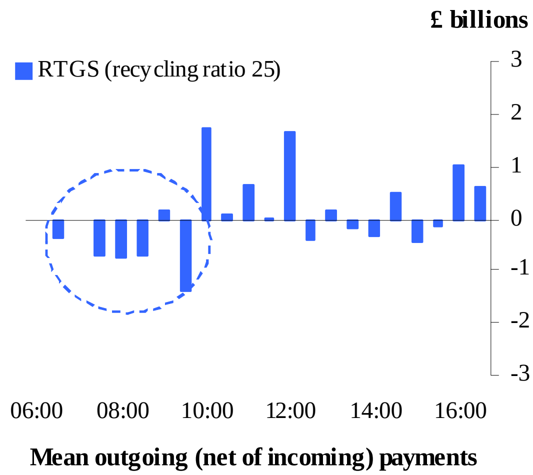


Figure 7.10 **Profile of payments sent under RRGS (net of incoming payments) for Bank A: mean value at half an hour intervals across the month**

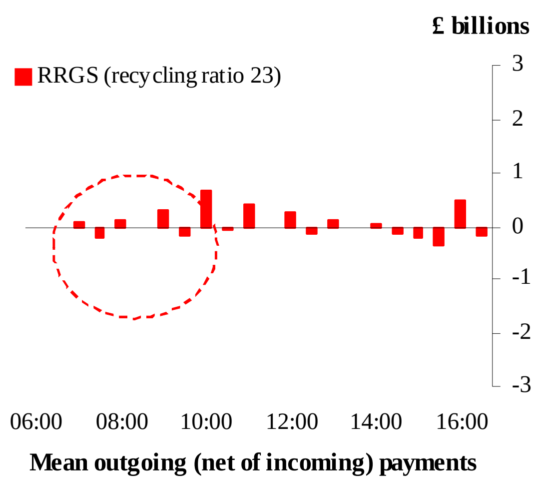


Figure 7.11 **Profile of payments sent under RTGS (net of incoming payments) for Bank B: mean value at half an hour intervals across the month**

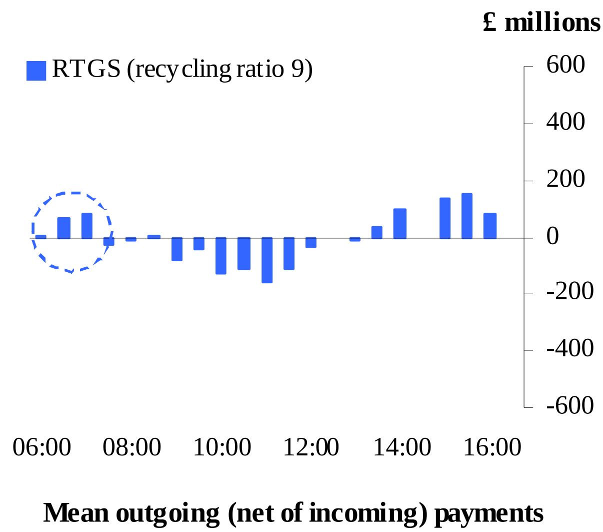
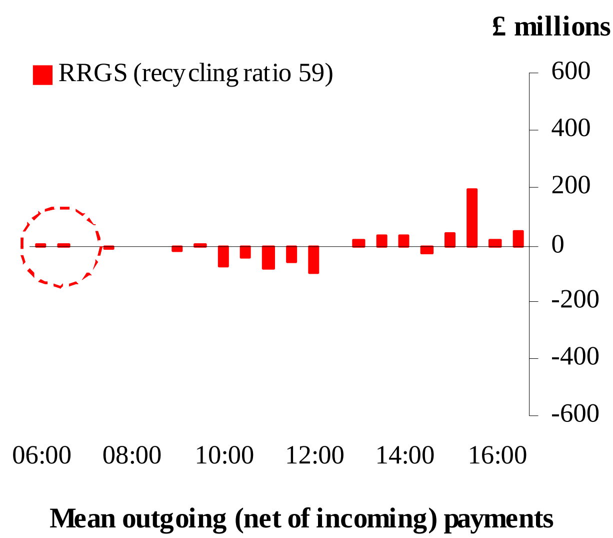


Figure 7.12 **Profile of payments sent under RRGS (net of incoming payments) for Bank B: mean value at half an hour intervals across the month**



Why do banks have markedly different recycling ratios under RTGS?

The answer lies partly in the intra-day profile of banks' payments. Becher et al (2007) suggest that structural differences in the underlying payment flows of banks may limit the extent to which payment timing can be managed so as to increase recycling ratios eg if certain banks (or their customers) routinely borrow in the overnight market, and others lend, the payment flows of the two groups will be correspondingly different.

Figures 7.13 and 7.14 below also help shed some light on the issue. Figure 7.13 shows that banks with high recycling ratios under RTGS see their liquidity demands increase (although they still remain the users with the highest recycling under RRGS), whereas those with the lowest ratios see the greatest savings. Figure 7.14 shows that larger banks typically have higher recycling ratios. This could be because a larger bank is likely to have more active links with the rest of the participants, increasing the probability of receiving incoming payments with which to recycle outgoing payments, especially in a system with extensive use of bilateral limits such as CHAPS. More fundamentally, the law of large numbers dictates that payment flows are more likely to be well balanced across the day where the volume of payments being made is greater.

Figure 7.13 **Impact of RTGS recycling ratio on banks' total liquidity requirements under RRGS**
Payments ≥ £1bn time-critical

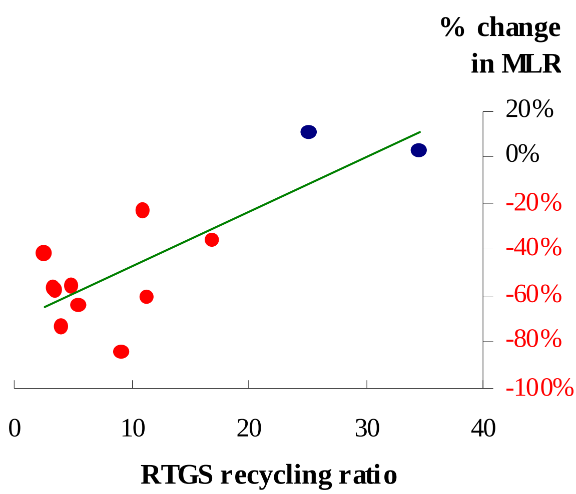
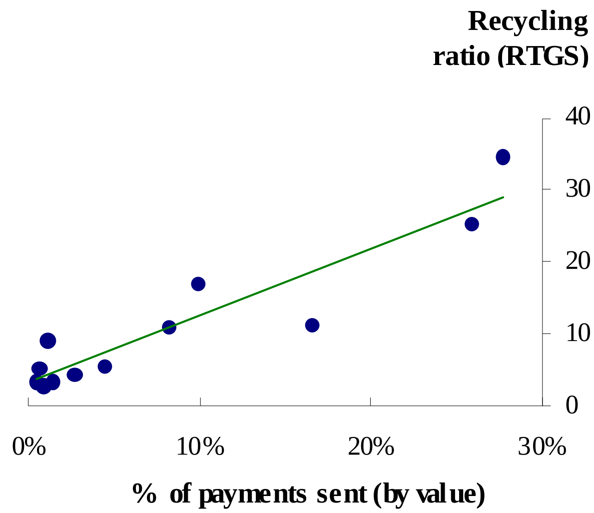


Figure 7.14 **Relationship between banks' total payment value sent and recycling ratio under RTGS**



7.4.3 Possible biases due to the use of bilateral monitoring

One source of payment delay in CHAPS, as discussed in Becher et al (2007), is that CHAPS banks typically monitor the sending behaviour of other members on a bilateral basis, in order to enable them to quickly detect if either: (i) another member has suffered an operational problem which is preventing them from sending; or (ii) another member is deliberately delaying payments in order to reduce their liquidity needs. Where banks observe an interruption in normal flow of payments from a counterparty for either reason they will typically stop sending to that counterparty until normal flow is resumed. Incentives for bilateral monitoring will remain under RRGS, but we are not able to capture the effects of the constraints placed on payment submission due to bilateral monitoring in our simulation methodology. This implies that our methodology may underestimate the amount of settlement delay introduced by RRGS by unrealistically relaxing one set of constraints on payment submission. More generally, this highlights that further work remains to be done carefully analysing the incentives created by the introduction of RRGS and how this would influence banks' payment sending behaviour.

7.4.4 Translating liquidity savings into cost savings

As discussed in Section 7.2, it is unrealistic to expect banks to post their exact daily minimum liquidity requirements into the system ex ante. Although the summary statistics we report are good indicators of the impact of RRGS functionality on banks' liquidity demands, their impact on banks' collateral posting decisions is not very clear. In some simulations, for example, RRGS reduces aggregate mean daily liquidity requirements across the month, but increases the aggregate of the maximum liquidity requirement across the entire month. So we need to assess how banks react to changes in the mean and standard deviation of their liquidity demands.

This is done by following the method employed in James and Willison (2004) of taking observed collateral posting decisions of CHAPS members and using regression analysis to estimate the influence of the mean and standard deviation of banks' max liquidity requirements on this decision. We use data for July-December 2006 to carry out a panel regression where the explanatory variables are the mean and the standard deviation of maximum collateral used on day t calculated over the 30 previous days in the sample; and the Libor / repo spread lagged by one day. In contrast to the methodology of James and Willison, a generalised least squares (GLS) estimator is used with correction made for serial correlation across time and within bank. The regression results are shown in Table 7.4.

Table 7.4 **Cross-sectional time-series GLS regression of collateral posted: all banks**
Including reserve account balances

	Coefficient	Standard error
Mean of maximum collateral used	0.24*	0.05
Standard deviation of maximum collateral used	0.98*	0.07
Libor / repo spread	0.00	0.00

Coefficients marked with * are statistically significant at 1%

All coefficients except the Libor / repo spread are in logs. The results could therefore be interpreted as: (i) a 1% increase in the mean liquidity requirement leads to a 24 basis point increase in collateral posting; and (ii) a 1% increase in the standard deviation leads to a 98 basis point increase in collateral posting.[19] The Libor / repo spread has a statistically insignificant impact on collateral posting, possibly because the true opportunity cost of posting collateral is low for banks subject to the UK Stock Liquidity Regime, who between them submit the majority of CHAPS payments by value.

We can now interpret our simulation results more clearly, by estimating the impact of RRGS on banks' collateral posting decisions using the above coefficients, our simulation results and the figure for mean collateral posting by a CHAPS bank in the 2nd half of 2006: £3.7 billion. This is shown in Table 7.5 below.

Table 7.5 **Estimating the impact of RRGS on banks' collateral posting decisions using James and Willison (2004)**

Criteria for time-criticality	Early submission	% Δ Liquidity requirement		Effect of Δ in mean & standard dev on collateral posting	Settlement delay (hh:mm)
		Mean	St dev		
≥£500mn	-	-10	+10	+0.3bn	+00:12
≥£1bn	-	-38	0	-£0.3bn	+00:37
Random 10%	-	-12	-10	-£0.5bn	+00:11
Random 3%	-	-37	-18	-£1.0bn	+00:25
≥£500mn	Random 50%	-15	+9	+0.2bn	-01:20
≥£1bn	Random 50%	-37	+5	-£0.1bn	-00:33
Random 10%	Random 50%	-14	-12	-£0.6bn	-01:15
Random 3%	Random 50%	-39	-18	-£1.0bn	-01:01
≥£1bn	Random 20%	-37	+2	-£0.3bn	+00:05
≥£1bn	Random 80%	-37	+2	-£0.3bn	-01:01

The key result in the above table is that the increase in volatility (as seen by the increase in standard deviation) causes mean savings under some simulations to disappear, and for two simulations results in higher collateral postings for the system as a whole: banks can post less collateral because their mean liquidity requirement is reduced, but they must post more to cover the extra volatility. As discussed in

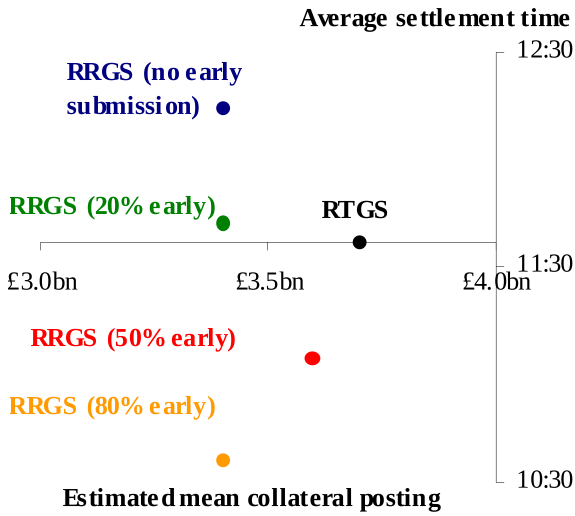
[19] The collateral posting decision is four times more sensitive to the volatility in daily liquidity needs compared with the mean liquidity need, suggesting that banks may also choose to post collateral for precautionary reasons.

Section 7.4.1, the increase in volatility is always observed with the value based time-criticality thresholds. Therefore under this metric the method of identifying time-critical payments does lead to significant differences in the costs faced by CHAPS banks.

7.4.5 Policy implications

Figure 7.15 shows the results of simulations where the ≥ £1bn time-criticality threshold has been used. The estimated mean level of collateral that needs to be posted is plotted on the x-axis against the mean time of settlement on the y-axis, bringing together sections 7.4.3 and 7.4.4. Unsurprisingly, RTGS requires the highest level of collateral. The blue observation is an RRGS simulation where historical payment data is taken as exogenous (which may be unrealistic). This shows banks face a trade-off: a decrease in collateral costs coupled with an increase in settlement delay. Altering banks' payment submission behaviour is shown to be able to improve this trade-off, where a significant fraction of payments are submitted early we see unequivocal improvements in aggregate welfare due to the introduction of RRGS, estimated collateral costs fall and mean time of settlement moves earlier in the day.

Figure 7.15 **Impact of different settlement and submission arrangements on mean collateral posting and average settlement time**
Payments ≥ £1bn time-critical
Average settlement times are measured on the 24h clock eg 11:30 means settlement in the late morning UK time



The results in Figure 7.15 give some weight to the idea that the introduction of hybrid functionality would be beneficial to CHAPS users. However, to make a firm policy recommendation about whether CHAPS should incorporate hybrid functionality further investigation would be needed to analyse the following

(i) What is a realistic assumption about banks' level of knowledge of their outgoing payments at the start of the day?[20]

(ii) What would be the development costs of incorporating RRGS into CHAPS?

(iii) Would banks trust the operational reliability of the central scheduler (anecdotal evidence suggests that this may be an issue)?

[20] Our dialogue with CHAPS users suggests a figure in the region of 20–50% may be realistic, although with significant variations seen across banks, primarily depending on the proportion of their payments that are associated with financial market flows.

(iv) What bias is introduced by our inability to incorporate the effect of bilateral monitoring on payment submission behaviour?

(v) What opportunity costs do banks face both from posting liquidity in CHAPS and from settlement delays?

If after considering these factors a decision were taken that RRGS would be beneficial, there are additional queuing functionalities that might bring further benefits if introduced in parallel. For example, a 'settlement no later than' option might encourage banks to submit even some 'semi time-critical' payments to the central queue (ie payments which for example, have to be settled in two hours, but can be queued until then for potential offsetting). If the settlement no later than time is reached and a payment has not been released for RRGS settlement, the queue places the payment into the RTGS stream for immediate settlement.

Gridlock resolution tools such as multilateral offsetting algorithms or the ability to reorder queues might also provide further optimisation, where large payments hold up conditional release of smaller payments in the queue due to the FIFO rule. However, our analysis suggests that only around 2–3% of all payments remain in the queue at end of day and need to be settled multilaterally net. This is the case even where 96% of payments by value are submitted to the central queue and suggests that the FIFO rule works well for CHAPS.

Finally, the fact that savings from RRGS are unevenly distributed, and some CHAPS users may even face cost increases, suggests that some coordination problems may be faced in getting general agreement on investment in RRGS. However, the fact that the smallest CHAPS banks get most benefit from RRGS is of interest to policy makers. One area of interest for the Bank of England over recent years has been in analysing the exposures created in CHAPS due to the highly tiered nature of its membership. As Harrison et al (2005) explains, a tiered structure where a large number of users make payments across the system indirectly through a correspondent bank has the potential to create credit exposures between correspondent banks and their customers. Therefore, to the extent that RRGS functionality could make direct membership of CHAPS more attractive to smaller banks, this might be favourably regarded by policy makers.

7.5 Conclusions

The analysis in this paper has modelled the impact of introducing a particular form of liquidity saving functionality, receipt reactive gross settlement, to the CHAPS system. Our simulations indicate that there could be benefits for CHAPS banks from the introduction of RRGS functionality. Two distinct improvements could potentially be captured: a reduction in the amount of liquidity that banks would have to post to settle their payments, and earlier settlement of those payments. It seems that a key determinant of the magnitude of the liquidity saving would be the proportion of payments that could be submitted to the receipt-reactive queue, while the improvement in settlement timing would depend on the proportion of payments whose submission is currently delayed.

Our results support the findings of Martin and McAndrews (2008) that RRGS can achieve a level of welfare that weakly dominates that achievable under RTGS. Our simulation approach expands on this result by allowing an analysis of the distribution of benefits across banks, demonstrating that typically the liquidity savings achieved would not be evenly shared across banks. The precise distribution of benefits is likely to be dependent on a range of factors relating to the structure of individual banks' payment flows under RTGS. In the case of CHAPS we find that banks with fewer payments typically face a proportionally higher liquidity need to settle these payments under a RTGS design. It is these banks who would benefit most from the introduction of a RRGS design. By contrast a subset of large banks who currently achieve high recycling ratios would see no savings and may even face a small increase in their liquidity needs.

We show that the value profile of time-critical payments can impact on collateral savings achievable under RRGS. In our simulations where the largest payments by value were time-critical, the volatility of liquidity demands faced by banks were typically higher than where values of time-critical payments were more evenly distributed in size. Volatility of liquidity needs is the key driver of banks' observed collateral posting behaviour.

Through analysis of a synthetic payment dataset we provide evidence that the key results discussed above are likely to hold in RTGS system more generally, and are not specific to the structure of CHAPS. We confirm that a key determinant of the impact of RRGS functionality is the liquidity recycling ratio that banks achieve under RTGS. We believe that this in turn is influenced by a number of factors including the number of direct members of the payment system, the volume and value of payments being processed in that system, the network topology of the system and the profile of payment values. This may shed light on the observation that the majority of payment systems which have already adopted hybrid designs typically had a large number of direct participants prior to adoption.[21]

Comparison of our results using the CHAPS and the synthetic datasets suggests that the existence of decentralised co-ordination mechanisms, such as bilateral monitoring, is already delivering some liquidity savings to CHAPS banks by reducing the volatility of their liquidity needs.

We have identified several promising avenues along which our work could be extended. When the Bank of Finland payment simulator is capable of replicating balance-reactive functionality then simulating the impact of its introduction to CHAPS would complement our analysis. More generally there is a wide range of liquidity saving functionality being employed in RTGS and hybrid systems at the present time, an extension of this methodology to a number of the leading approaches would be of interest.

A more detailed examination of the incentives banks face to submit payments to the queuing stream would be a worthwhile addition to our work, the benefits identified in our simulations are under the strict assumption that all banks will change their submission behaviour significantly to utilise RRGS. The differences in our findings depending on the method adopted to designate time-critical payments suggest there would be merit in exploring whether greater benefits could be achieved if banks co-ordinated on particular patterns of submission of time-critical payments early in the day to kick-start the release of queued payments. Finally, it would be interesting to explore the impact of RRGS under stressed circumstances ie where one or more CHAPS banks face an operational problem.

[21] Eg BI-REL in Italy (120 in September 2004) and RTGSplus in Germany (93 in December 2003): BIS (2005).

References

Bank for International Settlements (1997) **Real-time gross settlement systems.** Committee on Payment and Settlement Systems Publications No. 22.

Bank for International Settlements (2005) **New developments in large-value payment systems.** Committee on Payment and Settlement Systems Publications No. 67.

Bech, M – Preisig, C – Soramaki, K (2008) **Global trends in large-value payments.** Federal Reserve Bank of New York Economic Policy Review, September 2008, 59–81.

Beher, C – Galbiati, M – Tudela, M (2007) **The timing and funding of CHAPS Sterling payments.** Bank of England Working Paper. Economy Policy Review, Vol. 14, No. 2, 2008.

Harrison, S – Lasaosa, A – Tudela, M (2005) **Tiering in UK payment systems: credit risk implications.** Bank of England Financial Stability Review December 2005, 63–72.

James, K – Willison, M (2004) **Collateral posting decisions in CHAPS Sterling.** Bank of England Financial Stability Review December 2004, 99–104.

Johnson, K – McAndrews, J –Soramaki, K (2004) **Economising on liquidity with deferred settlement mechanisms.** Federal Reserve Bank of New York Economic Policy Review, December 2004, 51–72.

Leinonen, H – Soramaki, K (2003) **Simulating interbank payment and securities settlement mechanisms with the BoF-PSS2 simulator.** Bank of Finland Discussion Papers 23/2003.

Martin, A – McAndrews, J (2008) **An economic analysis of liquidity-saving mechanisms.** Federal Reserve Bank of New York Economic Policy Review, September 2008, 25–39.

McAndrews, J – Trundle, J (2001) **New payment system designs: causes and consequences.** Bank of England Financial Stability Review, December 2001, 127–136.

Millard, S – Polenghi, M (2004) **The relationship between the overnight interbank unsecured loan market and the CHAPS Sterling system.** Bank of England Quarterly Bulletin, Q1 2004, 42–47.

Willison, M (2004) **Real-time gross settlement and hybrid payment systems: a comparison.** Bank of England Working Paper No 252.

Appendix 1

Methodology for generating synthetic payments dataset

The basic method adopted to generate our artificial payments data was to use a Matlab program which implemented the following procedure

(i) Generate 4 columns (denoted A–D) of numbers of length n (the number of payments being generated) drawn from a log normal distribution. Generate an additional column (E) of length n drawn from a uniform distribution between 0 and 1.

(ii) Rescale and round to the nearest integer the numbers in A–C so that they contain integers between 1 and b, where b is the number of banks being simulated.

(iii) Rescale the numbers in D so that it contains numbers between 1 and m, where m is the largest possible payment being simulated. D represents the values of payments.

(iv) Compare matching rows of A and B. If $A_i = B_i$, and $E_i > 0.5$, replace B_i with a number taken from C. Continue taking consecutive numbers from C until $A_i \neq B_i$. Follow the same procedure, but replacing A_i with a number from C, if $E_i \leq 0.5$. Repeat process until all $A_i \neq B_i$. A denotes the sending bank, B denotes the receiving bank, the procedure is done to ensure no payments are sent and received by the same bank. Draws from E_i to decide whether the sending or receiving bank is replaced are intended to ensure that no bias is introduced to the probabilities of a particular bank being a sender or receiver of payments.

(v) Combine A, B and D with a column of payment times (T) which contains n entries evenly spaced between o (system open) and c (system close) to complete the dataset of artificial payments.

One extension to this method we used to make the dataset more realistic was to add some squaring-off payments at defined intervals through the day, which have the effect of ensuring that all banks end the day with zero balances. At the first two squaring-off points (after payments n/2 and 3n/4) each bank squares off half its accumulated balance, provided that balance has the same sign as its final balance at end of day. At end of day the remaining position is squared-off.

Rather than explicitly modelling squaring-off payments between banks we assume that all squaring-off payments are made with Bank b+1 which is introduced solely for this purpose. Throughout we quote results from a dataset including squaring-off payments.

We believe the introduction of squaring-off makes our dataset more realistic. In practice payment banks typically face a non-zero net payment flow on behalf of customers on each day, as payments sent are unlikely to perfectly match payments received. However, this net redemption is typically managed through overnight market transactions with counterparts, leaving most banks with final net balances that are close to zero.[22]

[22] The residual size of net flows is to some extent dictated by the monetary policy implementation regime being followed in a country. For example in the UK, since the introduction of reserve averaging in May 2006, banks do not face a requirement to completely square-off their net positions as this can be absorbed to some extent by changes to their overnight balances at the Bank of England. By contrast, prior to the introduction of reserve averaging, a strong incentive existed for banks to square-off to zero.

Chapter 8

Liquidity saving mechanisms and bank behaviour

Marco Galbiati – Kimmo Soramäki

8 Liquidity saving mechanisms and bank behaviour

Abstract

We investigate the benefits of liquidity saving mechanisms in interbank payment systems. We set up, simulate and compare two models, representing respectively a ‘vanilla’ payment system and a payment system with a liquidity saving mechanism.

In the first system, banks can route payments into a real-time gross payment stream (RTGS) or can queue them internally. In the second system, banks choose between the RTGS stream and a central liquidity saving mechanism (LSM), or central queue, that nets offsetting payments.

In both systems, at the start of the day, banks choose their opening liquidity balances for the RTGS stream. At the end of the day, they pay costs that depend on i) the chosen liquidity (liquidity costs) and ii) the delays experienced during the day (delay costs). As liquidity can be ‘recycled’, the delays suffered by any single bank depend on decisions of all the banks.

We compare the equilibrium choices in the two models with each other and with the choices of a benevolent planner. By so doing, we draw conclusions on the efficiency and desirability of the two systems.

8.1 Introduction

Interbank payment systems are used by banks to settle claims that arise from their trading with each other, or from customer transfers of funds from one bank or another. These systems form the backbone of the financial architecture, and their safety and efficiency are of great importance to the whole economy. The daily flow of payments in interbank payment systems generally amounts to 10% of a country’s annual gross domestic product (Bech et al, 2008). The main direct costs for banks in these systems (in addition to operations costs) are costs related to liquidity that is needed to settle the payments. On the other hand, banks may incur costs for delaying settlement.

Most large-value interbank payment systems use real-time gross settlement (RTGS) as the modality for settling payments. In RTGS, payments are settled individually and only if cover for their settlement is available. As a consequence, pure RTGS payment systems require large amounts of liquidity: if two banks have to make payments to each other, these transfers cannot be offset against each other: both banks must send the full payment to its counterparty. However, once a bank receives funds, it can ‘recycle’ them to execute its own payments.

This structure incentivizes free-riding: a bank may find it convenient to delay its payments (placing it in an internal queue), to wait for incoming funds and thus avoid the burden of acquiring expensive liquidity in the first place. There are three main reasons why such ‘waiting strategies’ in practice are limited, so that payment systems actually work: first, intervention by system controllers who typically penalize free riding behaviour, when detected. Second, peer pressure: the system’s participants themselves typically agree on common market practices and may punish non-cooperative behaviour. And third, delay costs: banks have an interest to make payments in a timely fashion; the cost of withholding a payment may eventually exceed the cost of acquiring the liquidity required for its execution, and so banks do not wait indefinitely.

However, it is well known that a certain volume of payments is internally queued for a while. While kept in the internal schedulers, these payments do not contribute to ‘recycling liquidity’, as they are kept out of the settlement process. A tempting idea is therefore to pool these payments in a central queue, to settle them more efficiently in a coordinated way; in particular, payments which offset each other could be settled without requiring any costly liquidity.[1]

Such central queues are called ‘liquidity saving mechanisms’ (LSMs) and systems employing them are generally termed hybrid systems. There are many varieties of hybrid systems but commonly they combine some form of net settlement for less urgent payments, while retaining the RTGS mode for more urgent ones.

Given the amounts of liquidity circulating in payment systems, the gains from hybrid features may be large. For example: to execute their payments, banks in the UK CHAPS system borrow from the Bank of England between 20 and 50 billion Pound Sterling on a daily basis, against high-quality collateral. And, the argument goes, this collateral

[1] It should be noted that if the mere submission to a central queue does not have legal implications in terms of settlement (ie payments are not settled until perfectly offset), then the settlement risk which led to the demise of end-of-day-netting systems, is not re-introduced. Hence, central queues with offsetting do not defeat the purpose of the gross payment modality.

may have more profitable uses elsewhere – for example, to collateralize securities clearing, interbank loans, or to generate income from securities lending. From another perspective: for a fixed amount of liquidity, if a payment system adopts an LSM, it may become more efficient, as the speed of settlement can be increased. Liquidity saving mechanisms have been on the agenda of policy makers for over a decade, and now many payment systems implement various central queuing facilities. While in 1999 hybrid systems accounted to 3% of the value of payments settled in industrialized countries, in 2005 their share was 32% (Bech et al, 2008). There is also a vast variety of them, differing in a number of dimensions. For example: how often should the controller look for payment cycles that can be netted? Should the LSM settle only perfectly netting cycles that require no liquidity at all, or should banks have the option of contributing any missing liquidity to accelerate the settlement process? Are submissions to the LSM irrevocable, or can banks retract payments from it, and when? Can individual banks monitor the central queue?

Liquidity saving mechanisms have been extensively studied via simulations, and existing systems have evaluated the effectiveness of their algorithms before implementation. Leinonen (2005 and 2007) provide collections of such investigations. Johnson et al (2004) proposes an innovative ‘receipt reactive’ settlement mechanism as an effective LSM. Both Guenzter et al (1998), and Shafransky and Doudkin (2006) develop approximate algorithms for solving the Bank Clearing Problem (the problem of selecting the largest subset of payments that can be settled with given liquidity), from an operations research perspective. Recently McAndrews and Martin (2008) developed theoretical models on liquidity saving mechanisms incorporating bank behaviour. Galbiati and Soramäki (2008), who study liquidity choices in an agent-based model of a payment system, forms the basis for this paper.

We argue that different LSMs give rise to different ‘games’ between system participants, who face differently shaped tradeoffs between liquidity and delay costs. This paper is a first exploration into these strategic aspects. Our very simple model includes the essential elements described above: payments, liquidity recycling, liquidity costs, internal queues, possible central queues, and delay costs.

We first model a benchmark case: a plain RTGS system with a queuing facility[2] where banks choose i) the amounts of liquidity to

[2] If a bank submits a payment to the system but does not have enough liquidity to settle it, the payment is placed in a queue and released immediately as liquidity becomes available from an incoming payment.

devote to settlement and ii) how many (and which) payments to hold in internal schedulers. Then, this case is compared to the case where an LSM is available. Here banks decide i) the amount of liquidity, as in the previous scenario, and ii) how many (and which) payments to submit to the LSM stream, In the LSM stream, payments are pooled and settled at zero liquidity cost by finding cycles of offsetting payments. Looking at these scenarios, we try to answer the following questions:

1. What is the outcome of a plain RTGS system where banks can internally queue payments? What are banks' equilibrium liquidity/queuing options when they have no mechanism to coordinate their actions? How do they compare to the choices of a 'benevolent planner' who maximizes social welfare?

2. By how much can our LSM theoretically reduce the needed liquidity and the delays? This is simply a question about the mechanical properties of the LSM.

3. What is the outcome of an RTGS system with an LSM? Is this efficient? How does this compare with the outcome without an LSM?

To anticipate some of the answers: 1) we find that in the system with only internal queues, individual banks underprovide liquidity and queue internally too much compared to what is socially optimal. This is due to the externalities in liquidity / queuing choices mentioned above (see eg Angelini, 1998). On point 2): when handled by a benevolent planner, our LSM may largely reduce liquidity needs. However, this is true only if the planner is not overly exigent regarding acceptable delays. Indeed, if delays must be reduced below a certain level, no payments can be queued and at that point having or not having LSM is irrelevant, as an LSM trades liquidity needs against delays. Finally on point 3), for an intermediate-range price of liquidity, an LSM may generate two different outcomes. One involves lower costs than without LSM – a 'good' equilibrium. However, the other outcome, under certain parameter values, entails higher costs than those in the absence of an LSM. Interestingly (and despite our initial intuition), the 'bad' equilibria are those with over-use of the central queue and higher liquidity usage.

These findings suggest clear policy implications: liquidity saving mechanisms are useful tools, but they require active management by the system's controller, or a coordination tool to ensure that banks adopt the low-cost equilibrium.

The paper is organized as follows: Section 8.2 describes the model; Section 8.3 solves it and presents the results. Section 8.4 concludes.

8.2 General framework

Our general framework is a simple model, adjusted in two different ways to describe the two systems that we compare. The model features N banks using a payment system. Banks make choices – to be illustrated later – that jointly determine system performance and hence the banks' costs or payoffs. The game-theoretic structure of the model is straightforward: a single simultaneous-move game, for which we find the Nash equilibria.

As described later, the model has an implicit time dimension. However, this only pertains to the settlement process, ie to the machinery used to derive the banks' payoffs. However, once the choices are simultaneously made, the expected-value payoffs are determined, so that there is no dynamic interaction between banks in a strategic sense. A main innovation of the paper is the way payoffs are determined: these are numerically generated by an algorithm which mimics a payment system in a fairly realistic way. We allow banks to exchange hundreds of payments over thousands of time-intervals, generating complex liquidity flows with 'queues', 'gridlocks' and 'cascades' (See Beyeler et al, 2007, for details on the physical dynamics of this process). We argue that this enhances realism by trying to incorporate the complex system's internal liquidity dynamics into the payoff function. Summing up, the model is a straightforward game-theoretic representation of a payment system whose complexity is encapsulated in the payoff function. And, such payoff function is computed via simulations.

8.2.1 Payment instruction arrival

Our model consists of N banks, which receive payment instructions (orders) from exogenous clients throughout a 'day'.

Each instruction is an order to pay one unit of liquidity to another bank. An instruction is therefore a triplet (i, j, u), where the first two indices represent the payer and payee and the third represents the payment's urgency (discussed below).

Payment instructions are randomly generated from time 0 (start of day) to time T (end of day) according to a Poisson process with given intensity. We assume that, at any time, any instruction (i, j, u) is equally likely to be generated. As a consequence, the payment system forms a complete and symmetric network, in a statistical sense: on average, the number of payments sent and received by a given bank vis-à-vis the other banks are equal. However, the days may differ: on one day a bank may be a net receiver, on others a net payer. And the order in which instructions arrive is also random.

The urgency parameter u, drawn from a uniform distribution U~[0,1], reflects the relative importance of settling the payment early: if payment r, with urgency u_r, is delayed by t time-intervals, it will cost the bank $u_r t$. Symmetry and completeness of the payment network are simplifying assumptions. However, they may be not be greatly unrealistic for many systems such as UK CHAPS, where specifically completeness is the typical case. Symmetry will be useful for technical reasons explained later on.

8.2.2 Payment settlement

A bank can route each payment into either of two streams: i) the RTGS stream or ii) a second stream. Payments submitted into the RTGS stream settle immediately upon submission, but only if the sender bank has enough liquidity. If the sender lacks sufficient liquidity, the payment is queued in RTGS and is released for settlement when the sender's liquidity balance is replenished by an incoming RTGS payment. Upon settlement, liquidity is transferred from payer to payee. For stream ii) we consider two cases, corresponding to two models.

The first model, without LSM, assumes that internal queues work in a very simple way: queued payments are withheld for the whole day and are submitted in gross terms to the RTGS stream at time T. A bank may want to queue a payment in order to reserve liquidity for more urgent payments. On the other hand, this model of internal queues is extremely simplified: in reality, banks delay payments only for a certain time, following more sophisticated rules. But our ultimate aim is to gauge the effects of introducing an LSM. To do so, it may makes sense to use as a benchmark an 'extreme' case where internal queues are managed in the simplest possible way.[3]

In the second model, banks have the option of routing payments into either RTGS or a liquidity saving mechanism (LSM). The LSM continuously nets payments on a multilateral basis; that is, it continuously searches and 'deletes' payments that form cycles of any size. To find offsetting payments, we use the Bech and Soramäki (2001) algorithm, which finds maximal cycles, under the constraint that each bank's payments are settled according to a strict order – here, by urgency. Payments settle in LSM only if they are exactly offsetting; hence the LSM requires no liquidity. However, any LSM payment that is unsettled at the end of the day, is moved into the RTGS stream and settled according to RTGS rules.

Our aim is to compare the benchmark system 'RTGS with internal queues' with the 'RTGS with LSM' system. The first system is a natural benchmark, because the option of internal queues is always available to banks. The second system is a specific example of a dual-stream system. Other LSMs could be considered, or other rules of interaction between streams could be considered. We choose the Bech-Soramäki algorithm for its simplicity and because it ensures an optimal outcome when payments are settled in a strict order, in our case by urgency. Finally, the possibility of combining RTGS to LSM with internal queues is ignored because, from the perspective of a single bank, LSM dominates internal queues: both mechanisms require very little intraday liquidity (only at end of day), but delays are slightly longer in internal queues (as there is a chance that a payment is settled before the end of the day with LSM). Hence, no rational bank in our setting would queue payments internally if LSM is available.

8.2.3 The game: choices and costs

At the start of the day each bank makes two choices: i) its opening intraday liquidity in the RTGS system $\lambda_i \in [0,\Lambda]$ and ii) an urgency threshold $\tau_i \in [0,1]$. Payment instructions with urgency greater than τ_i are settled in the RTGS system; the others are either queued internally or routed to LSM, depending on the model.[4] As the urgency parameter

[3] Note also that holding payments in an internal queue is always feasible for a bank.
[4] More complex routing rules are conceivable. We restrict attention to this for simplicity.

is drawn from U~[0,1], τ_i is also the percentage of payments that bank i queues internally or routes to LSM.

Once banks have chosen their opening intraday liquidity and urgency threshold, settlement of payments takes place mechanically: banks receive payment instructions and process them according to urgency.

Costs are defined as in Galbiati and Soramäki (2008). At the end of the day each bank pays a total cost, defined as the sum of a) the liquidity costs incurred in acquiring the opening intraday liquidity and b) the delay costs, which depend on the delays experienced during the day. Given a profile of choices $\sigma=\{\sigma_1,\sigma_2,...\sigma_N\}$ where $\sigma_i = (\lambda_i,\tau_i)$ is bank i's strategy, the costs borne by i are

$$\begin{aligned} C_i(\sigma) &= \alpha\lambda_i + D_i(\sigma) \\ &= \alpha\lambda_i + \sum_r u_r(t_r - t_r') \end{aligned} \tag{8.1}$$

where α is the price of liquidity and $(t_r - t_r')$ is the lag between reception and execution of payment r with urgency u_r. Delay costs thus increase linearly with payment urgency. The dependency of C_i on τ_i and all other σ_j's comes via the delays, which depend on the τ's and λ's of all banks in the system.

8.2.4 Equilibrium

The model has N players, actions λ_i and τ_i for each player, and costs/payoffs determined as described in the above section. We concentrate on the symmetric equilibria of this game, ie on those choices profiles $((\lambda_1,\tau_1),..(\lambda_i,\tau_i),..(\lambda_N,\tau_N))$ such that: i) all banks choose the same actions $((\lambda_i,\tau_i) = (\lambda_j,\tau_j)\forall i, j)$ and ii) each (λ_i,τ_i) is a best reply to others' choices.

By restricting attention to symmetric equilibria, we may miss equilibria where banks adopt different, albeit mutually optimal, choices. However, extra-model considerations suggest that such asymmetric equilibria (should they exist) would be unlikely to survive in reality. First, symmetry seems 'reasonable', as banks are homogeneous. Second, if a bank posted less liquidity than its partners, it might be seen to 'free-ride' and would be penalized in the long run. Finally, in the real world, banks do not know the choices of their counterparties; what they typically do know is some average indicator of the whole system, and this is what they play against. If N is large, all banks will face the same 'average opponent', and being identical, they will all choose the same best reply to that, which confirms that symmetric equilibria are the ones to concentrate on.[5]

8.3 Results

We illustrate first the mechanics of settlement; that is, we show how delay costs depend on banks' choices of liquidity and thresholds in the various streams. Then we illustrate the dependence of costs on banks' choices – this is the banks' payoff function. Finally, we identify and compare the corresponding equilibria.

All results are obtained by simulating the settlement process for different combinations of banks' choices.[6] As we look for symmetric Nash equilibria, we need run only simulations for each combination of 'my choices' vs 'others' choices where they do the same. This reduces the size of the parameter space to $([0,\Lambda]\times[0,t])^2$, from $([0,\Lambda]\times[0,t])^N$. Details on the numerical exploration of the delay and cost function are given in the Appendix. In most of what follows, we take the viewpoint of a single bank (I), facing the rest of the system (Them).

8.3.1 Settlement mechanics

Total delay costs D (see 8.1) accrue from delays in both RTGS[7] and in internal queues or the LSM mechanism. We show how these two sources of delays depend on the banks' choices.

---

[5] Equilibria where banks choose the same liquidity but different thresholds are unlikely for theoretical reasons. The more i uses the LSM, the more any other j should use it. The liquidity choices are different. Here, the substitutability effects may well induce asymmetric equilibrium behaviour: for example, low-l_i, high-l_j may be part of an equilibrium because, from i's viewpoint, j's liquidity is a substitute for i's own liquidity.

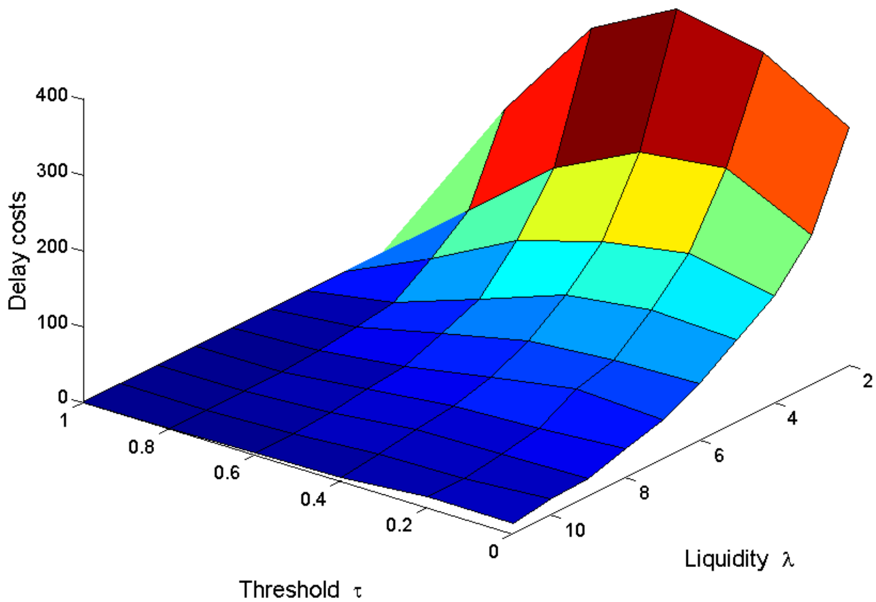
[6] As for the other parameters: the number of banks N is set at 15. The Poisson process generating payments is parameterized so that each bank sends on average 30 payments per day. We run ~200 days for each parameter combination.

[7] If a bank submits a payment but does not have sufficient liquidity.

8.3.1.1 RTGS delays

Figure 8.1 shows how delay costs in RTGS depend on λ and τ when all banks make the same choices (we choose this representation for clarity; in reality 'my' delays depend on four variables: my choices of λ and τ, and their choices of λ and τ).

Figure 8.1 **Delay costs in RTGS as a function of threshold and liquidity**



Obviously, delay costs are reduced by increasing liquidity (unless $\tau = 1$, because then no payment is actually directed into RTGS). And 'returns on liquidity' are decreasing, ie an additional unit of liquidity reduces delays more when liquidity is low than when it is high.

An increase in the threshold (ie less payments routed to RTGS) increases delay costs for low levels of τ – the more so, the less the available liquidity. This is probably due to the fact that, as low urgency payments are subtracted from RTGS, 'liquidity recycling' is disrupted. This effect is eventually balanced by the fact that fewer payments can be settled swiftly with less liquidity. Interestingly, liquidity has a stronger impact in reducing delays when not all payments are routed to RTGS ($\tau > 0$). Indeed, if all payments are routed to RTGS, liquidity is absorbed by less urgent payments too, so its 'returns' in terms of decreasing delay costs are reduced.

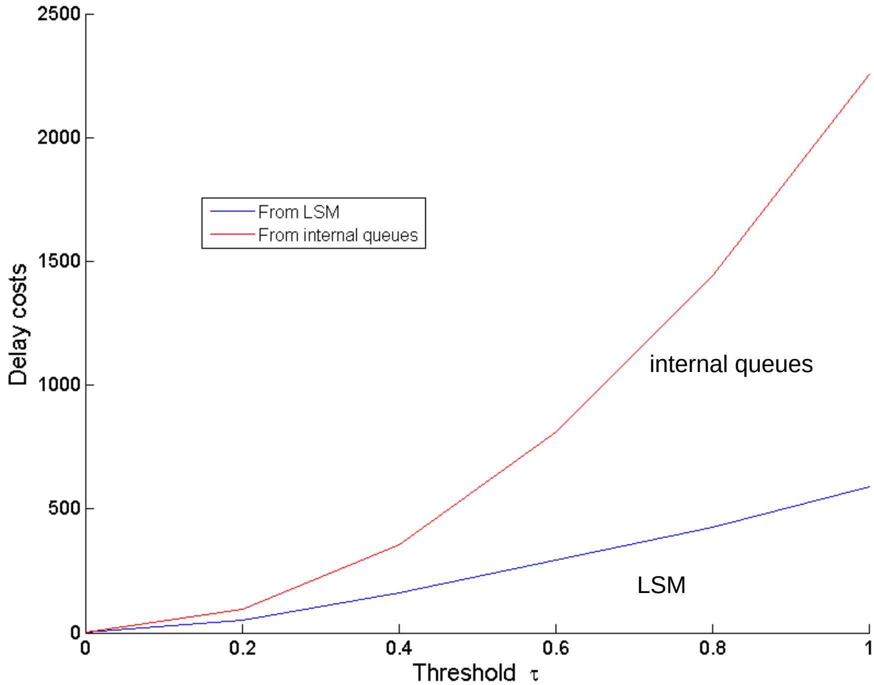
The relationship between τ and RTGS delay costs is generally non-monotonic: when liquidity is scarce, it is not convenient to route too many payments to RTGS: low-urgency payments may clog the system and cause more urgent ones to be unduly delayed. When liquidity is abundant, it is worthwhile to route all payments to RTGS, to minimize delays.

8.3.1.2 Second-stream delays

Delay costs in internal queues are simple. Obviously they are independent of λ, as internal queues consume no liquidity during the day.[8] On the other hand, internal queues' delay costs are a quadratic function of τ. Indeed, every payment settles at the end of the day, so the average time spent in the queue is half a day, ie T/2. The urgency of each payment is uniformly drawn from $[0,\tau]$, so it is $\tau/2$ on average. Hence, directing a volume of payments τ through internal queues produces delay costs totalling $(T/2)(\tau/2)\tau = (1/4)T\tau^2$.

[8] Only at the end of the day, are queued payments sent to RTGS and settled there. But, as they are added to the RTGS balance, the total amount of non-executed payments will equal the difference between incoming and outgoing payment orders. This is exogenous and so independent of banks' choices.

Figure 8.2 **Delay costs from internal queues and from LSM**

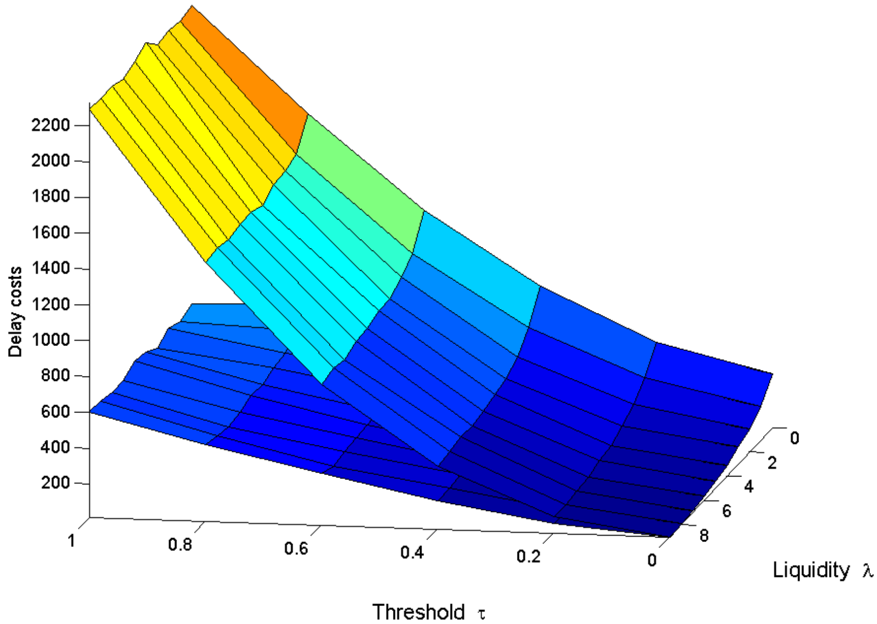


Delay costs in LSM are also independent of λ. More specifically total delays can be calculated as x·(τ/2)·τ, where x is the average time delayed, (τ/2) the average urgency and τ the volume routed to RTGS. Simulations show that total delays scale as α·τ, so one deduces that x~α·1/τ. In a sense, LSM displays increasing returns to scale with respect to processed volumes. The larger the pool of payments from which the algorithm can search for cycles, the more likely these cycles will be found. Delay costs generated by internal queues and by the LSM are compared in Figure 8.2.

8.3.1.3 Overall delays

Figure 8.3 shows how overall delay costs depend on λ and τ in the two systems (for illustrative purposes again we select the case of all banks making the same choice). Overall delay costs can be substantially reduced in the system with LSM (lower surface).

Figure 8.3 **Total delay costs for RTGS with internal queues (top surface) and RTGS with LSM (bottom surface)**



Figures 8.4 and 8.5 show the decomposition of delay costs into its two components: RTGS and second-stream (the four panels) are for increasing levels of liquidity.

Figure 8.4 **Total delay costs: RTGS with internal queues as a function of threshold. Each chart in the panel represents a given level of liquidity in the system**

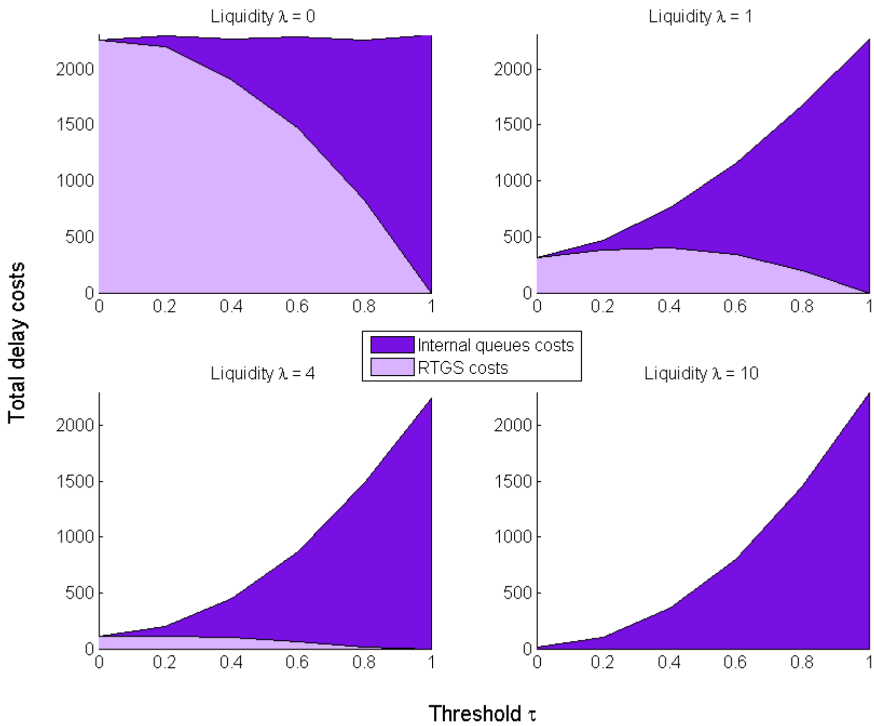
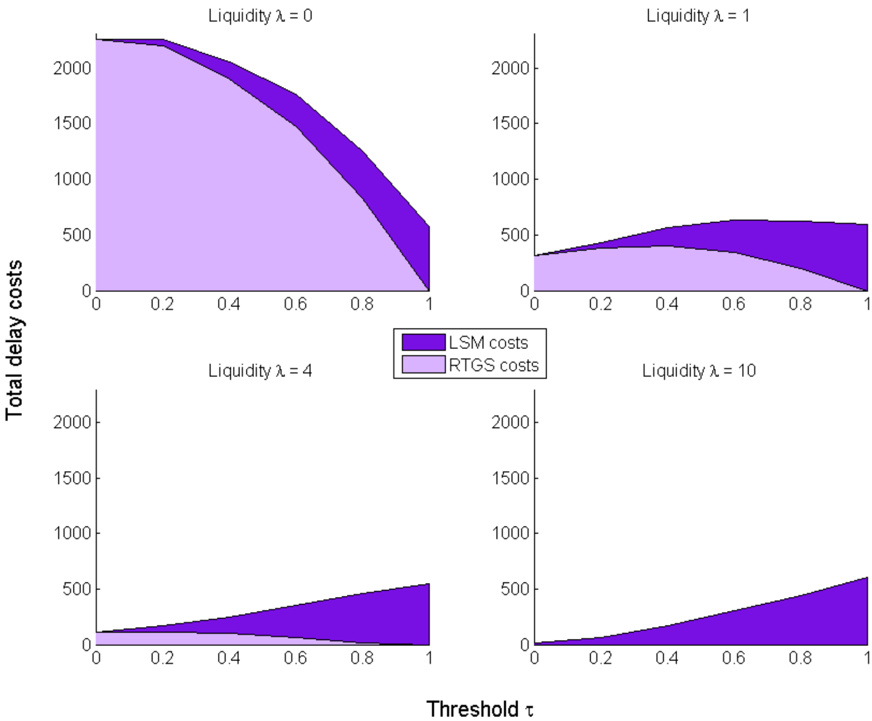


Figure 8.5 **Total delay costs: RTGS with LSM as a function of threshold. Each chart in the panel represents a given level of liquidity in the system**



8.3.2 Equilibria

The remainder of the paper looks at the equilibria reached by the banks in the two systems. A key parameter in the model is the price of liquidity, α, in Eq (8.1). This is arguably the variable over which central banks and policy makers have the greatest influence. Thus we look at how the equilibrium varies when the price of liquidity α changes. Because an accurate calibration of the model is beyond the scope of this paper, we let the parameter α vary in a range wide enough for the equilibria to span the whole strategy space – keeping the price of delays fixed.

8.3.2.1 RTGS with internal queues

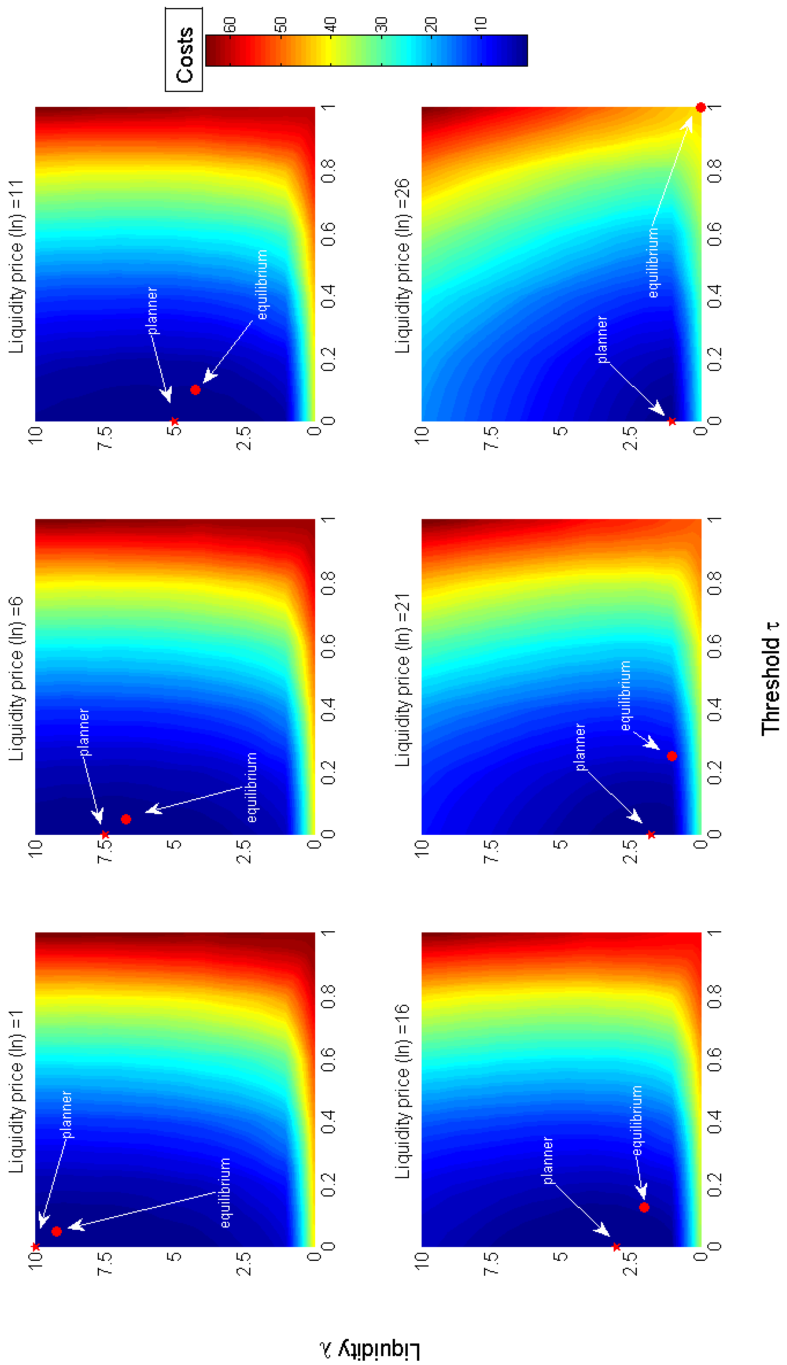
Figure 8.6 shows how the equilibrium in RTGS with LSM changes when the price of liquidity varies (increases, from left to right). Equilibrium choices are represented by dots, to be compared with the choices of a planner who minimizes the costs of the whole system. The background gradient shows system-wide costs. System-wide costs are minimized at the planner's choice; hence, the background gradient also shows how much higher the costs are at any (λ,τ) than at the planner's choice (dark blue = little higher, dark red = much higher).

Figure 8.6 shows that when the relative price of liquidity rises, banks post less liquidity and resort more to internal queues. More importantly, the equilibrium is inefficient: a cost-minimizing planner would provide more liquidity to the system and would delay less. Equilibrium costs are always more than 15% higher than the social optimum, reaching multiples thereof at high liquidity prices. Only when the liquidity price is extremely high, the equilibrium coincide with the planner's optimal choice, both being $\lambda=0$, $\tau=1$.

The reason for this inefficiency is explained by two externalities. On the one hand, a positive externality in liquidity provision: incoming payments to a bank can be recycled for making other payments, so liquidity is in a sense a common good, as in Angelini (1998) and Galbiati and Soramäki (2008). Due to this, equilibrium liquidity provision (λ) falls short of the social optimum. On the other hand, internal queues generate a negative externality: banks have an incentive to delay the less urgent payments and use liquidity for more urgent ones. But by doing so, they slow the beneficial liquidity recycling in RTGS, which harms other banks. Hence banks queue more than they should from a social perspective – ie τ exceeds the level that would be chosen by the planner.

It should be noted that the planner's choice of τ is dichotomous: either all payments are settled in RTGS, or they are all queued internally until the end of the day.

Figure 8.6 **Equilibria for RTGS with internal queues. Each chart in the panel represents costs and equilibria for a given price of liquidity (top-left: low, bottom-right high)**



8.3.2.2 RTGS with LSM

As with internal queues, LSM allows banks to reserve liquidity for urgent payments. However, internal queues merely postpone settlement until the end of the day, whereas our LSM allows for settlement as soon as offsetting cycles are found, without making use of liquidity. LSM therefore both reduces settlement delays and reduces liquidity demand.

Increased efficiency of the second stream induces banks to use LSM more intensely, with a reduction in costs. However, increase in τ also causes a reduction in RTGS volumes, which in turn causes this stream to loose in efficiency. Hence, there is a tradeoff between the efficiency levels of the two streams.[9] When 'played with' by individual banks, these effects produce unexpected outcomes, as we see next.

When liquidity costs change, the equilibria change essentially as in the RTGS with internal queues shown in Figure 8.6 (above). In particular: i) when the price of liquidity (α) rises, liquidity provision (λ) decreases and usage of LSM (τ) increases; ii) the equilibrium liquidity falls short of the social optimum and queued payments (τ) are in excess; iii) for very high relative prices of liquidity, both banks and the planner use exclusively the second stream – at which point the equilibrium is efficient; iv) the planner never uses both streams at the same time.

The main novelty with an LSM is that, for an intermediate range of liquidity costs, multiple equilibria emerge, as illustrated in Figure 8.7. Again, here we present the system's equilibria as the price of liquidity varies. For technical reasons,[10] we do not look at Nash equilibria, but at ε-Nash equilibria, ie at strategy profiles from which unilateral deviation yields a gain not exceeding a (small[11]) ε. These approximate equilibria are shown in Figure 8.7 as 'clouds' of dots; the socially optimal choice (planner's choice) is shown as a star.

[9] This is not the case with internal queues, where average delay times are independent of τ.

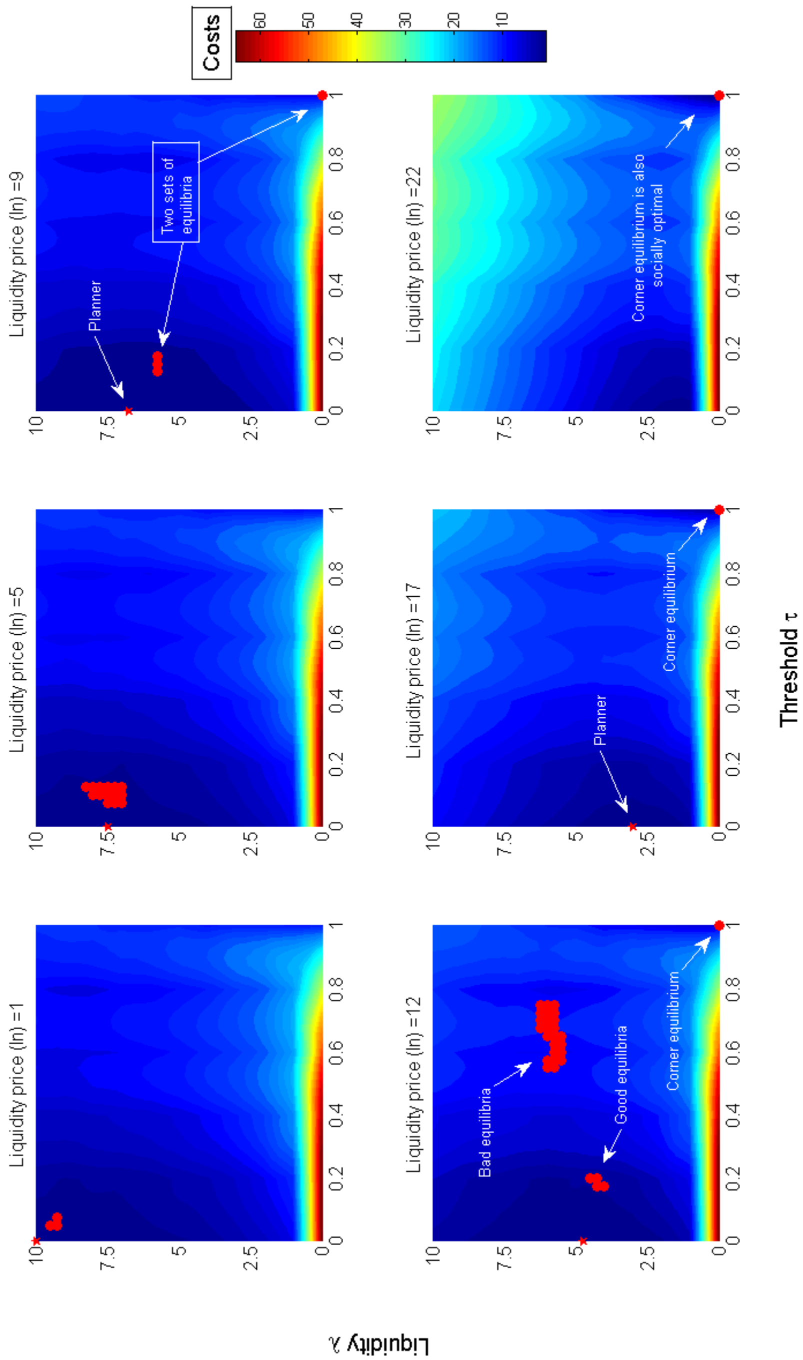
[10] The simulations show that, for certain levels of the price of liquidity, the payoffs become relatively insensitive to banks' liquidity-threshold choices. Because payoffs are numerically computed on a finite grid of choices, it is difficult to distinguish whether certain strategy profiles are exact equilibria or almost equilibria (ie deviations yield some gain).

[11] Figure 6 is obtained for $\varepsilon = 0.001$: a deviation does not improve payoffs by more than 0.1%.

The results are shown in Figure 8.7. When the price of liquidity exceeds a critical point, a corner equilibrium emerges, where banks acquire no liquidity and send all payments to LSM.

As the price of liquidity increases even further, the λ=0, τ=1 equilibrium persists, but other equilibria where banks use both streams and some liquidity emerge. Those with low λ display low costs – these are ‘good’ equilibria. The others are ‘bad’. Apart from the corner equilibrium, the bad equilibria are somewhat paradoxical: they feature higher costs, higher liquidity usage (λ) and higher LSM usage (τ). The existence of such equilibria is probably explained as follows. LSM features economies of scale (see Section 8.3.1.2), so that high usage of it may be self-sustaining. But, as mentioned at the start of this section, over-use of LSM is detrimental to the RTGS stream – which may then require higher amounts of liquidity.

Figure 8.7 **Equilibria for RTGS with LSM. Each chart in the panel represents costs and equilibria for a given price of liquidity (top-left: low, bottom-right high)**

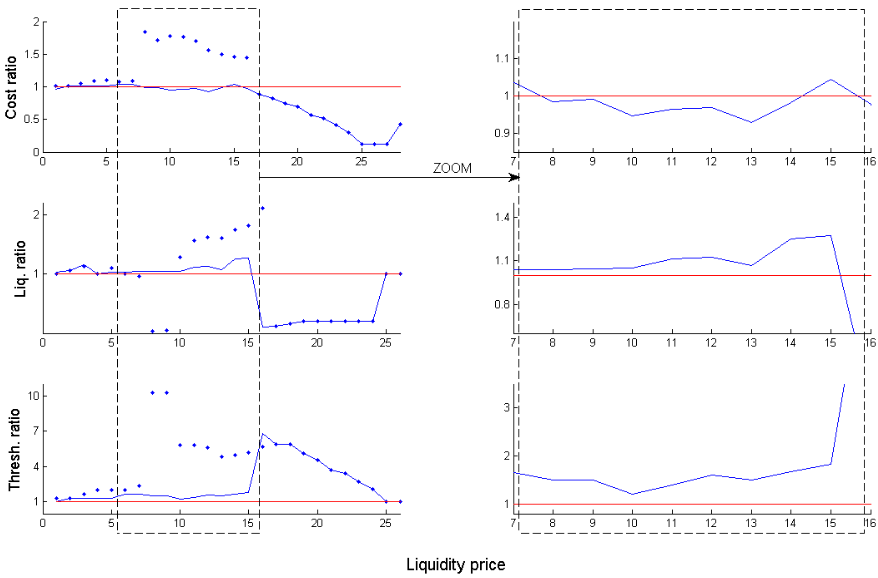


8.3.2.3 Comparison of the two systems

The key comparison of this paper is between two systems: RTGS with internal queues and RTGS with LSM. With LSM we have ‘clouds’ of ε-equilibria; hence, for each cloud, we pick average values of costs, liquidity and thresholds. A ‘bad’ cloud is then determined by the highest average costs, and a ‘good’ one by the lowest average costs.

The blue lines in Figure 8.8 show the equilibrium costs attained with an LSM, normalized by the corresponding costs obtained without LSM – solid line for ‘good cloud’ values and dots for ‘bad cloud’ values. The chart on the right shows a sub-area of the left chart for intermediate liquidity price values. Eg. when liquidity price $\alpha=13$, the good LSM equilibria are about 5% cheaper than the equilibrium with internal queues. Savings become more sizable at higher prices of liquidity.

Figure 8.8 **Ratio of cost, liquidity and threshold between the good and bad equilibria varying cost of liquidity**



8.4 Conclusions

This paper compares two stylized payment systems. In both of them, banks can queue non-urgent payments, to reserve liquidity for the urgent ones. In the first system, queued payments are held 'internally', and are settled at the end of the day. In the second system, queued payments are placed in a central queue (LSM) and are settled throughout the day as cycles form. As expected, central queuing (LSM) is more efficient than decentralized queuing (as in a system with internal queues). However, we find that the 'mechanical' advantages of an LSM can be nullified by strategic behaviour: as the model shows, there exist 'bad' equilibria with high liquidity usage, intense use of the LSM, and yet costs that exceed those obtained in a system without LSM. These findings suggest that liquidity saving mechanisms are useful tools, but they may need some coordination device, to ensure that banks arrive at a 'good' equilibrium.

A necessary caveat is that this paper considers one specific liquidity saving mechanism and compares it to another specific (rather extreme) model of internal queues. Other LSMs, perhaps associated with different settlement rules, may yield different outcomes.

References

Angelini, P (1998) **An analysis of competitive externalities in gross settlement systems.** Journal of Banking and Finance, Vol. 22, 1–18.

Bech, M – Preisig, C – Soramäki, K (2008) **Global trends in large value payment systems.** Federal Reserve Bank of New York Economic Policy Review, Vol. 14, No. 2.

Bech, M – Soramäki, K (2002) **Liquidity, gridlocks and bank failures in large value payment systems.** E-money and Payment Systems Review, Central Banking Publications, London.

Beyeler, W – Bech, M – Glass, R – Soramäki, K (2007) **Congestion and cascades in payment systems.** Physica A, Vol. 384, Issue 2, 693–718.

Buckle, S – Campbell, E (2003) **Settlement bank behaviour and throughput rules in an RTGS payment system with collateralised intraday credit.** Bank of England Working Paper No. 209.

Galbiati, M – Soramäki, K (2008) **An agent-based model of payment systems.** Bank of England Working Paper No. 352.

Güntzer, M M – Jungnickel, D – Leclerc, M (1998) **Efficient algorithms for the clearing of interbank payments.** European Journal of Operational Research 106, 212–219.

Johnson, K **–** McAndrews, J J **–** Soramäki, K (2004) **Economizing on liquidity with deferred settlement mechanisms.** Federal Reserve Bank of New York Economic Policy Review 10, No. 3, 51–72, 2005.

Leinonen, H (2005) (ed) **Liquidity, risks and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies, E: 31.

Leinonen, H (2007) (ed) **Simulation studies of liquidity needs, risks and efficiency in payment networks.** Bank of Finland Studies, E:39.

Martin, A – McAndrews, J J (2008) **Liquidity-saving mechanisms.** Journal of Monetary Economics, Vol. 55(3), 554–567.

Shafransky, Y M – Doudkin, A A (2006) **An optimization algorithm for the clearing of interbank payments.** European Journal of Operational Research 171(3): 743–749.

Willison, M (2005) **Real-Time Gross Settlement and hybrid payments systems: a comparison.** Bank of England Working Paper No. 252.

Appendix

To compute the payoff function of bank i (Eq. 8.1), we need to find the delays experienced by i when the rest of the system chooses $\{(\lambda_j\tau_j)\}$_$\{j\neq i\}$. As mentioned in the main text, we can treat the 'rest of the system' as one player and assign to it symmetric action profiles $(\lambda_j\tau_j)=\{(\lambda_1\tau_1),(\lambda_2\tau_2),..\}$ such that $(\lambda_1\tau_1) = (\lambda_2\tau_2) =$.... This greatly reduces the action profiles to explore, because now the delays $D_i((\lambda_i\tau_i),(\lambda_i\tau_j))$ are a function of 4 variables only. We compute them as follow.

We run simulations for a restricted number of 2-player action profiles. In particular, we simulate the settlement process for λ taking on all integers in [0,10], and τ any number in [0,0.2,0.4,..1]. That is, we compute $11_2 = 121$ values of the delay function, for just as many action profiles. To do so, because payment orders arrive in a random order, we need to simulate at least 200 'days' for each action profile to obtain a reliable estimate of the 'average day'. Hence, we simulate $200*11_2 = 24'200$ days in total.

Yet, 11 choices for each bank are not enough to obtain 'smooth' results: when computing the equilibria, undesired artefacts emerge. Hence, we numerically smooth out and interpolate the delay function $D_i((\lambda_i\tau_i),(\lambda_i\tau_j))$ on a refined grid, a 4-dimensional cube with $41_4 = 2'825'761$ points, which correspond to banks choosing λ in [0,10] in steps of 0.25 (41 liquidity levels) and τ in [0,1] in steps of 0.025 (41 threshold levels). This is the delay function $D_i((\lambda_i\tau_i),(\lambda_i\tau_j)) = D(\sigma)$. Adding liquidity costs, we obtain the cost, or payoff function defined in Eq. 8.1). Using such payoff function, equilibria are computed – numerically, of course.

Chapter 9

Participants' internal intraday limits in large value payment systems

Matti Hellqvist

9 Participants’ internal intraday limits in large value payment systems

Abstract

Participants in large value payment systems can manage their intraday liquidity and risk related to intermediation of payments with procedures which control the flow of transactions. One alternative is to implement a cap for the net outflow of liquidity towards a given counterparty based on the values of received and sent payments during the day. These caps are called bilateral intraday counterparty limits. Such limits can cause negative externalities if delays are increased in the processing of payments through the payment system.

Participants can implement bilateral counterparty limits in their internal systems. This paper proposes a simple simulation based methodology, which can be used to implicitly estimate the level of such internal limits from transaction data of the payment system. The method is used for data from BoF-RTGS, the Finnish large value payment system, from years 2002–2007. The estimated levels are used in linear regression, where the observed values are explained with external variables related to counterparty risk. Secondly the estimated values are compared to overnight loan positions between the same participants.

No stable levels of bilateral counterparty limits were observed in the study from Finnish data. In the regression analysis there was no clear connection between counterparty risk related variables or overnight loan positions and the estimated limits. Several possible improvements in the methodology are also discussed.

9.1 Introduction

In a modern market economy, payment systems can be considered a basic part of the society’s infrastructure because all economic transactions in the real economy are eventually settled in some manner in the payment systems. Large value payment systems (LVPS) constitute the core network of the payment systems. The most common LVPS structure is real-time gross settlement (RTGS), where transactions are settled in central bank money and are final immediately upon settlement.

To enable settlement of transactions, banks in an RTGS system must have liquidity in their central bank account during the day. The word liquidity refers here to funds that enable immediate settlement of transactions in central bank money. The three possible sources of such liquidity are credit from the central banks provided preferably against collateral, interbank markets and incoming payments from counterparties. Studies of intraday liquidity management in RTGS systems have shown that banks may have an incentive to delay transactions if intraday liquidity is costly. This may cause inefficiencies in the overall level of the respective payment system because of negative externalities.[1]

One possible tool for payment system participants' intraday liquidity management is to implement limits for intraday counterparty positions. Such limits restrict the outflow of payments and liquidity provided to counterparties, unless there are also incoming payments. Bilateral limits can be explicitly defined in some LVPSs, as in TARGET2 in the euro area and LVTS in Canada.[2] However, if the system does not provide facilities for defining bilateral limits on liquidity flows, banks can implement limits in their internal systems. Based on anecdotal evidence, many banks have such systems in place.

Another line of research, which shares some aspects with the current paper, comprises the studies of operational incidents in large value payment systems and their impacts on participants' behaviour. One example is Bedford et al (2005), which reports simulations of operational incidents in UK's Chaps system. In that study the reactions of other system participants were explicitly modelled as internal stop sending rules, where payments to the failed participant are suspended or postponed. Such behaviour requires close monitoring of payment streams and facilities similar to internal bilateral intraday limits. Another related example is by Klee (2007), who reported an analysis of the existence of operational incidents in Fedwire and their impact on the money market.

This study proposes a methodology for identifying internal intraday counterparty limits of banks empirically from payment system data. The method is tested with transaction data for the Finnish

---

[1] See eg Angelini (1998), Bech-Garratt (2002) and also Hellqvist (2005).

[2] Limits in Canada's LVTS system are due to the difference in the basic structure of the system, which is based on credit flows between system participants instead of immediate transfers of liquidity as in an RTGS system. See Dingle (1998) or Arjani and McVanel (2006).

LVPS (BoF-RTGS) from the years 2002–2007. For estimating the internal limits from payment flows, the system setup is replicated via the Bank of Finland Payment and settlement system simulator.[3]

The validity of the method and received estimates for internal bilateral counterparty limits are tested using two approaches. Both are based on assumption that intraday liquidity management reflects counterparty risk management: if a counterparty is more highly trusted, a larger intraday exposure vis-à-vis that counterparty would be allowed. In the first test the estimated levels of counterparty limits are included in a regression where the independent variables are external market based measures of counterparty risk. In the second approach, the magnitude of intraday positions is compared to overnight loan exposures. These are identified from payment system data via a commonly used methodology (Furfine, 1999). The magnitude of intraday exposures should correlate with the overnight loans if there are some counterparty credit lines in place and these are successfully captured in the estimates.

The proposed methodology can be useful for oversight of payment systems. Some possibilities for its application are discussed, such as monitoring the changes in intraday liquidity usage or identifying free riding participants who may adversely affect the efficiency of the payment systems.

The paper is structured as follows. Section 9.2 presents the estimation of internal limits and the data used in this phase; section 9.3 describes the tests of validity of the estimated limits; and section 9.4 concludes.

9.2 Estimation of participants' internal bilateral limits

The proposed method of estimating participants' internal bilateral limits is based on transaction data from a payment system. For each transaction, the timing, the involved participants and the value must be known. When the processing logic of the system is known in detail, the flow of payments and resulting participants' balances can be reconstructed and recorded at any time.

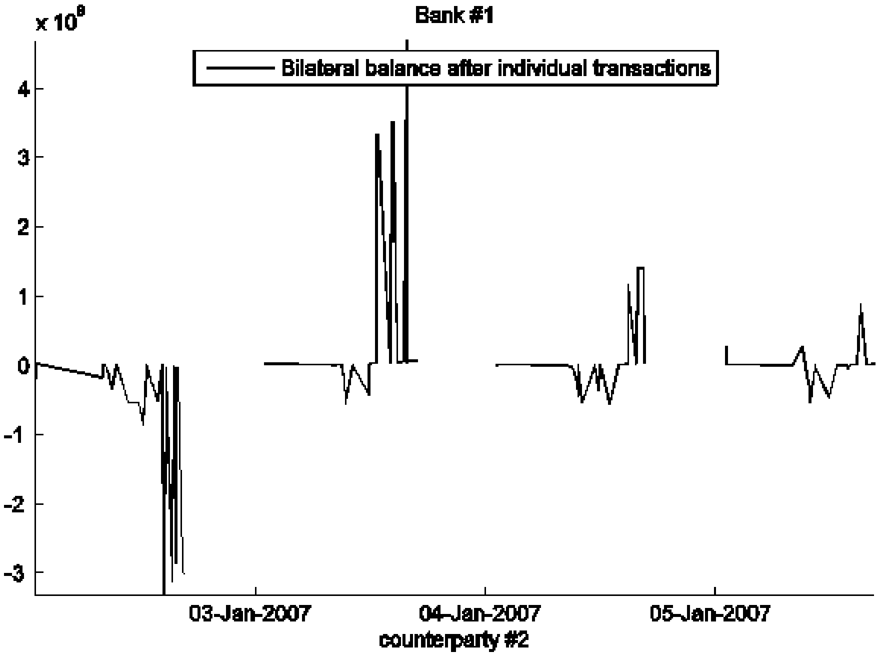
For this study, the Bank of Finland Payment and settlement system simulator was used to replicate the processing pattern of the payment

[3] See www.bof.fi/sc/bof-pss/ for more details on the simulator.

system. For estimating the internal limits, the bilateral limits feature of BoF-PSS2 is used. In a simulation with this feature, bilateral intraday positions are automatically calculated and it is possible to define explicit bilateral sending limits, which cannot be exceeded during the simulated settlement process. In the input data for the simulator, the bilateral limits are set at sufficient high values to make them non-binding constraints. When transaction flows and bookings are reproduced identically to the actual history, the observed bilateral positions remain within the assumed – but unknown – binding constraints, ie the bilateral limits that were in place in the participants' internal systems. This enables implicit estimation of the bilateral limits set by the participants.

When the internal limits are measured via the proposed approach, some specific characteristics have to be considered. Participants' bilateral balances can be viewed as a stochastic process, where settled transactions change the balance. In artificial setups, transactions values in large value payment systems have been approximated with lognormal distribution (eg Baksys and Sakalauskas, 2006). When this is combined with normal payment processing procedures, such as the first in first out rule (FIFO), large transactions can block the process even if the balance is far from the actual limit. This can make identification of existing limits challenging. On the other hand, banks may be able to reorder payment instructions or allow smaller payment to bypass larger ones in their internal queues if there is not enough liquidity to settle a large payment without violating the counterparty limits. If one assumes there are internal limits and thus willingness to delay at least some transactions, the possibility of reordering the payments and bypassing the FIFO rule is a natural addition. This could make it easier to observe the limits if such are used in real systems. One observed realisation of the bilateral positions between two participants from the data under study is shown in Figure 9.1 below.

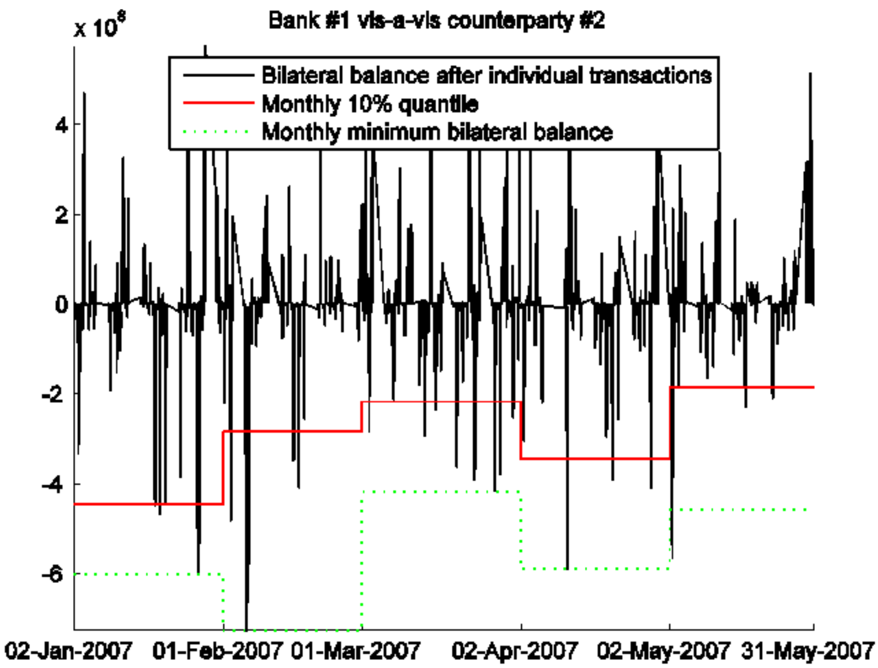
Figure 9.1 **Example of bilateral intraday positions between two banks after each individual transaction on the four fist days of year 2007. Balance is plotted only for the system's open hours**



The simplest approach to estimate the bilateral limit is to select a period of time and use the minimum bilateral position between the given participants as the estimated limit. Discussions with practitioners imply that while there are systems with counterparty limits in place, such limits can also be rather flexibly changed or bypassed by prioritized transactions. This would cause the simple minima of bilateral balance to become a biased estimator for the actual counterparty limit. As an alternative, quantile estimates of the limit were calculated. In quantile estimation, most cases are assumed to stay within a given fixed limit and a small number of transactions is allowed to cause bilateral positions which violate the limit. The quantile can be calculated so that only eg 1% or 5% of all transactions exceed the limit. This decreases the impact of possible outliers or manual changes in the estimated limit. The method is a rather coarse heuristic approximation, but since the objective is to capture the magnitudes of limits, not their exact values, this is perhaps not a big problem. Figure 9.2 shows the bilateral position over a longer period and monthly minimum positions as well as monthly 10% quantile estimates of the bilateral limit.

An alternative method using penalty functions was also tested. In that approach, the limit was searched with maximization of the limit value and a linear or quadratic penalty term was included in the objective function for each bilateral position violating the limit. The heuristic quantile method was however used instead of penalty functions because it was found to be a simpler, robust and equally accurate way to estimate the magnitudes of assumed internal limits.

Figure 9.2 **Bilateral positions for two banks during five months, monthly minimum and 10% quantile estimates of bilateral limit**



The system under study in this paper is BoF-RTGS, the LVPS operated[4] by Bank of Finland. Data consisted of all transactions from March 2002 to end-December 2007, which came to 1494 individual days and 70 months. The estimation of counterparty limits was done for daily and in monthly frequencies at several quantile levels. In all,

[4] BoF-RTGS was closed in February 2008 when Finland joined the TARGET2 system.

24 participants were included. The total number of accounts in the data set is higher, but some special participants and account classes were excluded (eg Bank of Finland or technical accounts). The data set includes all the largest banks operating in Finland.

9.3 Validity of limit estimates

Estimated counterparty limits were tested in two ways which are related to the assumption of bilateral limits as a tool or reflection of counterparty risk management. These were a regression analysis versus external variables associated with counterparty riskiness and a comparison of intraday bilateral positions with the magnitudes of overnight loans. Both tests are presented in detail below.

9.3.1 Regression for counterparty risk measures

If participants' internal bilateral limits are used for limiting intraday counterparty risk, the level of observed limits should correlate with external variables describing the soundness or riskiness of each counterparty. The assumption was tested via regression analysis with the estimated internal intraday limit as the dependent variable. The regressions were performed using both ordinary least squares and stepwise regression. Before any regressions were run, Dickey-Fuller unit root tests were performed on the estimated internal intraday limit values. Only the results of such cases where the estimated limit data passed this unit root test at the 0.05 confidence level are reported below.

All regressions were computed for both daily and monthly level data sets. Four different quantile levels for the dependent variable were tested[5] for the daily level data and seven for the monthly data. With multiple quantile estimates, it was possible to check for differences in how well the different levels are explained by the independent variables.

Two external variables were used to approximate counterparty riskiness. One is a distance-to-default indicator (DD) calculated from counterparties' equity prices and balance sheet data based on the

[5] At the daily level, minima and 1%, 2.5% and 5% quantiles for the bilateral position were studied. At the monthly level, minima and 0.5%, 1%, 2.5%, 5%, 7.5% and 10% were used correspondingly.

option pricing model of Merton (1974). Distance to default or similar indicators are commonly used as a credit risk management tool. Data for the five publicly listed banks[6] included in the current dataset were based on an earlier study at the Bank of Finland (Lehto, 2005).

The second independent variable was the spread between collateralized and uncollateralized three month Euribor and Eurepo interbank rates. The interest rate spread (IS) is used as a general measure of the confidence between market participants. This data are available as from March 2002, when the Eurepo interest rate was launched, which defined the data period for the whole study.

The third independent variable was the moving average over the previous 20 days[7] of daily total incoming transaction volumes (TV) for a given counterparty. It can also be considered as a partial risk management measure, since two banks that engage in many mutual transactions should each have good knowledge of the other as a counterparty and mutual trust. Even more, this variable should describe the intraday liquidity management since, if you expect to receive a high value of payments from a particular counterparty, you will likely also send payments to that party.

Three variations of the two first independent variables were tested: the simple face value of DD and IS values; the interaction of DD or IS with the number of payments sent to a given participant; and the interaction with the logarithm of the number of corresponding payments sent. The last two options are motivated by the relatively small transaction volumes. With very few transactions, it is likely that the limit will not be reached – even if such is in place. As a result, the estimated value of a limit can become smaller than the actual limit in the real system. With a larger number of transactions, it is more likely that the limit becomes effective and thus the importance of variables related to counterparty riskiness should increase. However, if there are very many transactions and the same limit is assumed to be binding, an increase in the number of transactions should not have such a large impact, which motivates the use of logarithm for transaction volumes. The moving average of transaction volumes over the 20 last days was used for the interaction variables.

All independent variables were available for five of the 24 counterparties. Thus it was possible to test 115 bilateral pairs of

[6] Oko, Nordea, Sampo, Ålandsbanken and eQ.

[7] Time windows of 1,5,10 and 20 days were tested for calculation of moving averages. Moving average with 20 day period was most often a significant independent variable.

counterparties.[8] For each individual pair, an independent OLS regression was run if the limit series, ie the dependent variable, passed the unit root test. Thus the IS data were the same in each individual regression as it is general and market level variable. The DD data were specific to the receiving participant and thus shared by all regressions with that given counterparty. Only the dependent variable data and the coefficients or independent variables derived from the transaction flows were unique for each pair of banks.

There were no significant differences in the regression outcomes for dependent variables estimated with different quantile levels of daily data. The reason for this is in the relatively low daily transaction volumes in the system, which further on causes only very small differences in the daily limit data at the different quantile levels used. All the results below are from regressions with daily minima as the dependent variable. DD and IS variables were included in the final daily model multiplied by the transaction volume logarithms. A summary of the regression results for daily level data is presented in Table 9.1 below.

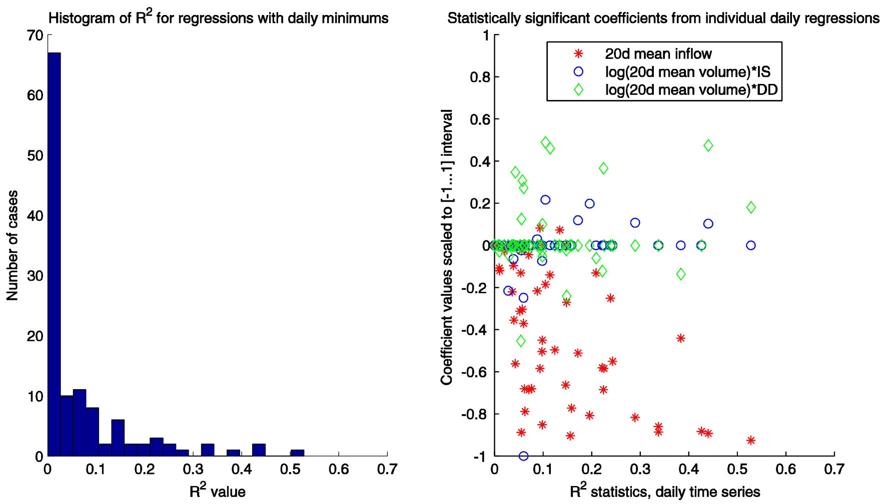
Table 9.1 **Descriptive values of one selected regression with daily level data of bilateral limit estimates and counterparty risk measures**

Different independent variables in regressions with daily time series	Number of times when significant, (before 1-root filter)	Sign of the coefficient
Log(20-day mean volume)*Distance to default (DD)	14 (16)	?
Log(20-day mean volume)*Interest rate spread (IS)	24 (31)	?
20-day mean value, incoming payments (TV)	50 (58)	negative
Nr significant independent variables		
cases filtered based on unit root test	7%	
0	43%	
1	25%	
2	22%	
3	3%	
Highest individual R^2 value	0,528	

[8] 115 pairs were formed from 5 participants against 24 counterparties, excluding the meaningless (bank x, bank x) pairs.

Based on the results, in 50% of the cases one or more of the independent variables was statistically significant. Figure 9.3 below presents the statistically significant estimated coefficients from the regressions. The figure also shows the histogram of R^2 statistics from all of the regressions. Transaction value was the only independent variable with good explanatory power for the estimated limits. The negative sign of the coefficient accords with intuition, since the limits have negative values in the data. The results for the DD and IS variables were mixed, and the sign of the impact was not clear.

Figure 9.3 **R^2 statistics and estimated coefficient for all individual bank pairs from regressions with daily level data**



With monthly data, a larger scale of possible quantile levels was tried, and there were some differences in estimated limits and regression results with different quantiles. However, when the R^2 values for regressions with different quantile levels were compared with statistical testing,[9] there was no evidence of a difference between the distributions of R^2 values. This holds for all pair-wise tests between the quantile levels, even when only the nonzero R^2 values were used. Thus no such quantile level was found that would fit better than others

[9] Nonparametric Wilcoxon rank sum test was used to test the hypothesis of whether the distributions of R^2 values observed in regressions with different dependent variables (see footnote 5) are the same.

in the set of independent variables, and all the results are again reported from regressions based on minimum bilateral balances data.

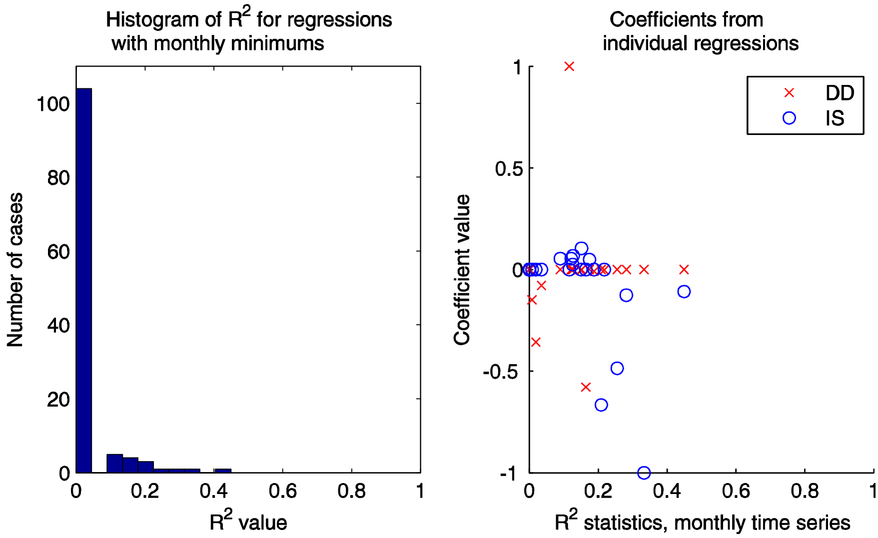
The first significant difference between monthly and daily data was found in the unit root tests. In fact, 73% of possible pairs of banks (84 out of 115) were filtered out in this phase. For the rest of the cases, the results were similar to those for the daily regressions. The regression results for the plain IS and DD are summarized in Table 9.2 below. The histogram of R^2 values and values of statistically significant coefficients are also displayed in Figure 9.4.

Table 9.2 **Descriptive values of one selected regression with monthly level data of bilateral limit estimates and counterparty risk measures**

Different independent variables	Number of times when significant (before 1-root filter)	Sign of coefficient
DD	19 (19)	?
IS	19 (33)	– (?)
Nr significant independent variables	Nr of cases	
cases filtered based on unit root test	73%	
0	10%	
1	0%	
2	17%	
Highest individual R^2 value	0,449%	

The results do not clearly indicate the signs of the coefficients, at least for the DD variable. For IS, it was found that in all cases where the variables have some explanatory power the sign was negative. The estimated limits were in the data with negative values, and so it might be that if the interest rate spread (IS) has an impact on internal counterparty limits, a wider spread between collateralized and uncollateralized interbank credit could make the banks more eager to allow large intraday positions.

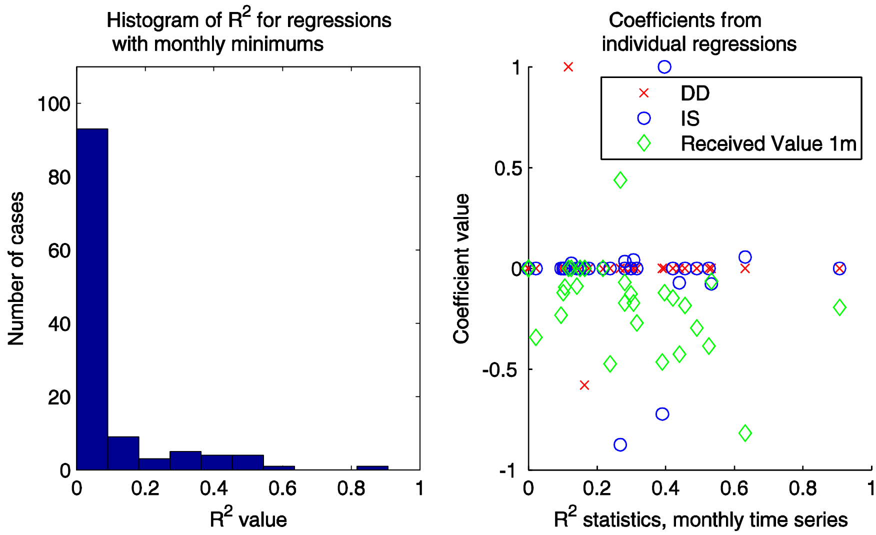
Figure 9.4 **R^2-statistics and coefficients from a set of regressions with monthly level data**



If the value of the payment inflow is included as the third independent variable, the variations in the dependent variable are explained slightly better. Some individual cases with high R^2 values or almost perfect fit are also found. However, the correlation of monthly transaction volume with DD values is relatively high.[10] Moreover, when the payment inflow is included, it becomes the only independent variable for which the sign of the coefficient could be reliably observed (see Figure 9.5 below).

[10] The correlation between monthly DD and incoming liquidity was 0.76.

Figure 9.5 **Regression statistics for monthly level data with value of incoming payments included**



9.3.2 Comparison of intraday positions with overnight lending volumes

As an alternative test for the internal counterparty limits, their magnitudes are compared to overnight loan positions, which can also be estimated from the payment system data. If the intraday bilateral limits are actually used as a limit for the counterparty risk, the level of allowed intraday position should also accord with the acceptable maximum for the overnight positions. This approach avoids some problems of regression between estimated bilateral limits and counterparty risk measures, since the real credit limit against each counterparty, including all liabilities, should be almost the same during a given day and at the end of the same day.

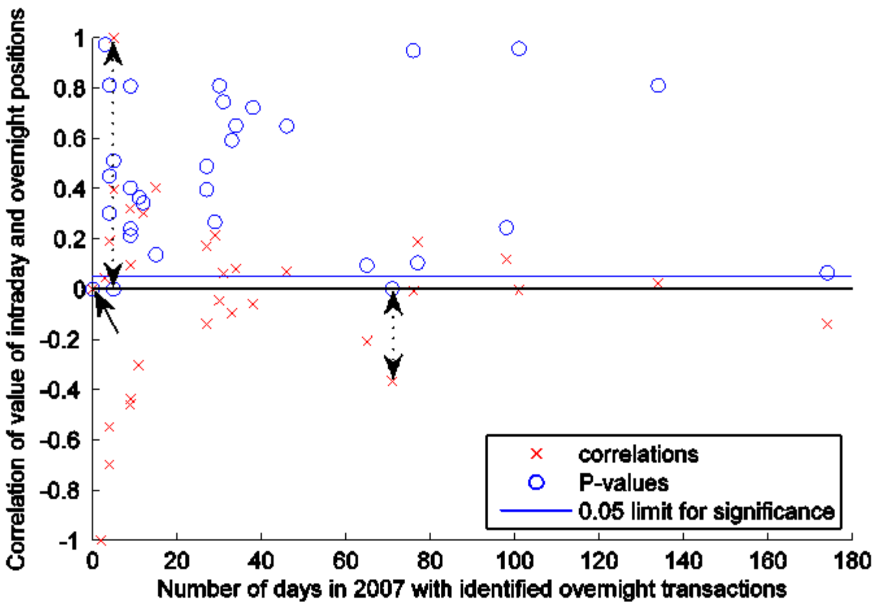
A frequently used method for identifying overnight loans from payment system data has been presented by Furfine (99). The idea is to locate transactions between banks A and B where the value is a large round number and where there is a transaction in the opposite direction on the next possible date with principal and interest closely matching the interbank rate. In this paper, multiples of 10 000 Euros were searched and at most a 50bsp difference versus the Eonia overnight interest rate was allowed.

For this comparison, overnight loans were sought only from year 2007 data. There were in all 8607 transaction pairs located, corresponding to 2.6% of the total transaction volume. Of these transactions, 2022 were between 12 of the 24 banks studied in the regressions. The rest were primarily cross border overnight loan transactions.

Correlations were calculated for the daily data of estimated intraday bilateral limit and total value of overnight loans for each pair of banks using Pearson's linear correlation coefficient. For each correlation coefficient, also a P-value was calculated from Student's t distribution based on the number of observations. The values of correlations and corresponding P-values are shown in Figure 9.6 below against the number of days from which the value of overnight loans could be identified for the pair of banks in question.

Figure 9.6 **Correlation of daily values of overnight loans and value of minimum intraday bilateral positions calculated independently for each pair of banks.**

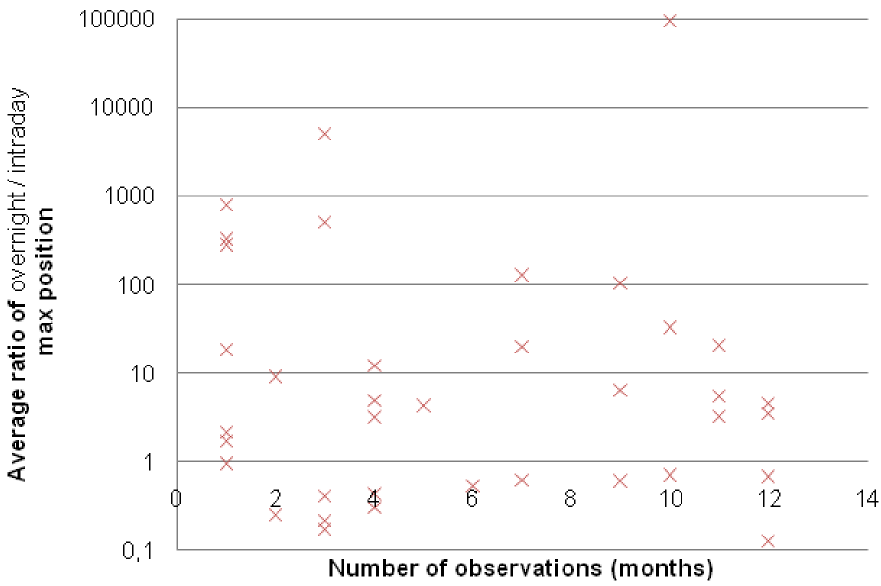
X-axis gives the number of observations (days) with some identified overnight loans for the given pair of banks.Three cases for which the P-value indicates statistically significant correlation are marked with arrows.



No clear sign for the correlation can be seen since the three cases out of 56 with P-value smaller than 0.05 are likely to be just random noise. In addition, the presented analysis did not include filters for unit roots as in the regression results presented earlier.

The overall magnitude of intraday and overnight positions can also be compared. Minimum intraday bilateral positions for each month were compared to maximum daily overnight positions from the same month and for each bilateral pair of banks. Below in Figure 9.7 the average of these ratios from the year-2007 data are shown as a function of the number of months for which the ratio was available for a given pair of banks.

Figure 9.7 **Average ratios of the magnitude of monthly minima of intraday position and maximum overnight position between individual pairs of banks. Relation is plotted on logarithmic scale as a function of number of observations for each pair of banks**



The ratios are very large. The smallest average value is 0.1, which implies intraday exposures ten times the overnight exposures against the same counterparty. The largest individual ratio, averaged over ten months, was overnight exposures: one hundred thousand times the smallest intraday position. The wide variations are also visible within the ratios of individual banks; the same bank may have a very small ratio of overnight vs intraday positions against one counterparty and a high one against another.

9.4 Summary and discussion

Transactions from the Finnish large value payment system, BoF-RTGS, were used to identify assumed internal bilateral limits for the intraday counterparty positions. The data consisted of all transactions settled in BoF-RTGS between March 2002 and December 2007.

Regression analysis was performed between the estimated intraday position limits and independent variables related to counterparty risk and the value or volume of transaction flows. The explanatory power of the counterparty risk-related independent variables was weak. If internal bilateral limits are used by Finnish banks, it is more likely they are used for intraday liquidity management. There is clearly no common practice among all participants to adjust or set bilateral intraday limits for counterparties based on counterparty risk management measures.

One finding of this study is that the difference between the uncollateralized and collateralized interbank interest rate may be positively correlated with the size of allowed intraday positions. Thus higher uncertainty could occur together with wider intraday limits. This could be understood as reluctance to reveal any type of liquidity problems to counterparties if the level of confidence between market participants is low. This observation is related to the recent period of liquidity turbulence, since before August 2007 there was practically no significant difference between the Euribor and Eurepo interest rates. However, a further study is required to confirm the observation.

A second test for the estimated levels of bilateral intraday limits was performed by comparing them with the overnight positions between the same participants. These were similarly identified from the payment data using a common methodology. No statistically significant correlation could be found in the current data between the value of overnight positions and the value of estimated bilateral limits.

The overall level of intraday positions was noticed to vary greatly as compared to overnight loan positions. The wide range of ratios can be explained on the basis of at least two factors. The first possibility is that the structure of overnight lending markets and intraday payment flows is very different and thus those banks which are used for overnight credit are not the same as those to whom most payments are sent during the day. This alternative is not dependent on the existence of internal limits. The second alternative is that banks have limits for their intraday positions after all, and with these limits the intraday positions against some counterparties are kept small enough for the high value of ratios to emerge. The latter case would mean that limits were mainly for intraday liquidity management, which would be possible if intraday and overnight activities are taking place between the same counterparties. Analysis and comparison of the network structure of the intraday payment flows and overnight exposures would be needed to determine the reason for the observed ratios.

There are some known shortcomings in the current analysis. Only transactions taking place in one LVPS are considered. In real life, banks may have intraday positions against each other in many other systems such as securities settlement systems, foreign exchange settlement, retail payment systems and private large value payment systems or even other central banks' RTGS systems. If there is an in-house system for intraday positions and limits, it would ideally cover all these positions. In this study, only the data from the Finnish large value payment system were used with the assumption that it should contain the relevant part of the value of intraday positions between the market participants who are active in Finland.

In real systems, the banks may or may not include the operations performed on behalf of their customers in their bilateral positions controlled by the bilateral limits. If limits are used for counterparty risk management, executing payments for the bank's customers should not perhaps be limited. However, if the limits are primarily for intraday liquidity management, the original reason for a payment may be irrelevant.

One basic assumption in this study is that bilateral limits are stable and fixed at least for each day or even for each month. Based on anecdotal evidence, this may not be the case. Instead, the limits may be quite flexibly changed and, for example, may as a rule be initially low and then gradually relaxed. The approach of this study did allow outlier transactions to bypass the rigid limits, since quantile estimates were tested, which increases flexibility. Also the assumption of the more stable or fixed level of intraday limit can be understood as the final level to which the real, and possibly variable, limit can be extended. Such 'backbone' limit should indeed be based on more stable variables such as perceptions of counterparty riskiness or overall intraday liquidity capacity.

Based on the data characteristics, time series analysis or co-integration analysis could be more suitable approaches than the ordinary least squares regression used here. Also, in this study all the pair-wise regressions were managed as one homogeneous group. In the real world, the practices and thus the results for individual institutions are likely to be different. Filtering or grouping the estimation and tests based on the sender bank would allow for analysis of the practices of individual banks against all of their counterparties. These analyses are left for future studies.

Transaction volumes in large value payment system are not very high. This can make the estimation of bilateral limits more difficult; a larger number of observations would more likely reveal the level of the estimated implicit limit. This idea could also be used to test the estimated limits. First, it should be assumed that there is a given binding limit for the bilateral position between banks A and B set by the bank A. The bilateral position after each transaction can be considered a random walk, which will hit the limit more likely when there are more steps, ie a higher volume. Thus the observed value of the limit should increase when the transaction volume is low and increases. After some level, which is the actual real limit, the increase in volume no longer has an impact on the expected value of the observed limit.

This approach can be seen as a test for the use of bilateral limits as an intraday liquidity management tool because it is based on the process of liquidity flows and no external variables are needed. However, the test has to be performed for each participant pair separately and for some pairs there are no very high volume days available. Also, a rather large sample size is needed and thus it will not be easy to track changes in the limit levels.

New developments in LVPS's allow more advanced liquidity management within the centralized payment system. As an example, TARGET2 includes the possibility to explicitly define bilateral sending limits. In such a case the proposed indirect method is not necessary and the efficiency of the system or rationale for the levels of the limits could be analyzed directly. Explicit limits also allow development and testing of the proposed indirect method, which could be useful in other systems without explicit limits. All these tasks are left for future studies.

Transaction data from payment systems is quantitative data with high accuracy and high frequency that can be used for identification of positions and actions of participants in payment systems. Although this study did not find a clear rationale for the observed intraday positions, which would be based on counterparty risk measures, the methodology can provide valuable insight for oversight purposes.

Comparison of participants' intraday and overnight positions should reveal the sizes of counterparty positions in different time scales or maturities. Also the analysis can reveal whether there any implicit limits in the systems or whether there are changes in the levels of these limits due, for example, to changes in market conditions.

Analysis of intraday behaviour of payment system participants would also reveal which participants are repeatedly using more than average incoming payments to fund their intraday liquidity needs. If such free riding is observed, some policies for safeguarding the efficiency of liquidity circulation in the system can be considered. Such possibilities might be changes in the settlement logic to render the process more efficient and reduce the incentive to delay payments.

Finally, the findings presented in this paper will depend on the market practices or market specific features of any particular system. Thus the results from one system should not be generalized without critical assessment or own estimations of other setups.

References

Angelini, P (1998) **Analysis of competitive externalities in gross settlement systems.** Journal of Banking and Finance 22, 1–18.

Arjani, N – McVanel, D (2006) **A Primer on Canada's Large-Value Transfer System.** Bank of Canada Web Document, www.bankofcanada.ca/en/financial/lvts_neville.pdf.

Baksys, D – Sakalauskas, L (2006) **Modelling of interbank payments.** Technological and Economic development of Economy, Vol. XII, No. 4, 269–275.

Bech, M – Garratt, R (2002) **The intraday liquidity management game.** Journal of Economic Theory 109, 198–219.

Bedford, P – Millard, S – Yang, J (2005) **Analysing the impact of operational failures in large-value payment systems: a simulation approach.** In Leinonen, H (ed. 2005), Liquidity, risks and speed in payment and settlement systems – a simulation approach. Bank of Finland Studies, E31.

Hellqvist, M (2005) **Intraday liquidity management in gross settlement system as a coordination game.** Financial Markets and Statistics department Working paper 06/2005.

Klee, E (2007) **Operational problems and aggregate uncertainty in the federal funds market.** Mimeo, (Presented at the Bank of England & European Central Bank conference on Payments and Monetary and Financial Stability, November 12–13, 2007)

Lehto, J (2005) **Distance-toDefault – An Indicator of Bank Fragility.** Bank of Finland, Financial Markets and Statistics department Working paper 05/2005.

Merton, R (1974) **On the pricing of corporate debt: the risk structure of interest rates.** Journal of Finance, 29, 449–470.

Chapter 10

The sterling unsecured loan market during 2006–2008: insights from network topology

Anne Wetherilt – Peter Zimmerman – Kimmo Soramäki

10 The sterling unsecured loan market during 2006–2008: insights from network topology

Abstract

We visualise the unsecured overnight market in the UK as a network of relationships and examine how it has changed over the period of market turmoil. Using established network techniques, we find strong evidence of the existence of a ‘core’ of the most connected banks. We find that this core has become more important during the crisis, and that the widened reserve target bands have allowed banks to exercise more discretion in forming relationships. However, when for a short while the core banks appear risky, correspondents prefer to diversify and reduce their reliance on the core.

10.1 Introduction

Financial innovation in general, and securitisation in particular, feature heavily in common explanations of the sources of the recent market turmoil. Yet, in many ways this period has seen a number of classic disintermediation forces at work.[1] In this paper, we aim to illustrate this disintermediation by examining data for the UK unsecured interbank market throughout 2006–2008. Using a unique set of individual trades in the UK CHAPS interbank payment system, we employ topology methods to assess how the network of lending relationships between individual banks changed.

Although overnight unsecured sterling activity is not necessarily lower during the crisis compared to pre-crisis levels, we do find a reduction in the number of bilateral relationships, suggesting that the network has become characterised by a core of relatively few banks regarded as bearing less risk to their counterparties. Although this core exists pre-crisis, we suggest that it becomes more important during the crisis phases as participants become more aware of counterparty risk. We also observe that when for a short while the core banks appear

[1] See eg Borio (2008) and Brunnermeier (2008).

more risky, correspondents prefer to diversify and reduce their reliance on the core.

Our results are consistent with many of the classic features of disintermediation explained by the theoretical banking literature. First, theoretical models tell us that interbank lending activity frequently falls during crises, as banks reduce their credit exposures. Flannery (1996) shows that financial intermediaries reduce their lending when they become more concerned about their ability to assess their counterparties' credit worthiness. Moreover, during a crisis, individual banks in their model are no longer able to diversify their loan portfolio as markets become less liquid. Consequently, they abstain from lending activity altogether. Freixas and Jorge (2007) model the impact of an aggregate liquidity shock and show how severe liquidity shortages may arise. Together with asymmetric information, this causes liquidity in the interbank market to flow towards the most liquid banks, at the expense of the less liquid ones. Likewise, Acharya et al (2008) show that during a liquidity crisis, liquidity-poor banks will find access to the interbank market greatly reduced, as liquidity-rich banks exert their market power and charge higher rates. This forces the former to exit the interbank market and rely on asset sales instead.

Taken together, the theoretical literature suggests that interbank markets may cease to function efficiently when concerns about credit worthiness increase and banks are hit by aggregate liquidity shocks. The result is an overall reduction in interbank activity, often accompanied by a reallocation of flows away from weaker banks.

There is a second mechanism at work. The banking literature also demonstrates that, when faced with unexpected shocks, banks may want to build up their own liquidity reserves. Freixas et al (2000) consider a situation where lenders withdraw from the market because they are uncertain about their own ability to borrow in the future.

Allen et al (2008) show that banks reduce their interbank lending when there is uncertainty about the overall demand for liquidity in the banking system. In this model too, banks cease to use the interbank market and start hoarding liquidity. Caballero and Krishnamurthy (2008) demonstrate that when Knightean uncertainty (ie uncertainty about future states of the world) increases, financial intermediaries are inclined to assume the worst-case scenario and hoard liquidity. This type of hoarding typically occurs when unexpected events happen and coincides with investors reducing their risk exposures. Hence, this second line of research points to an overall reduction of lending activity which takes the form of a flight to higher-quality, and hence more liquid assets.

In a recent paper, Ashcraft et al (2008) challenge this view and argue that faced with increased uncertainty about intraday liquidity shocks, banks may hoard liquidity in the early part of the day. But later in the day, as their payment obligations become clearer, they are more willing to lend their excess reserves overnight. This would explain both rises in overnight lending activity and large intraday variations in overnight rates. Furthermore, they show how smaller banks, who typically face greater liquidity constraints, build up larger intraday reserves and tend to be net lenders to the larger banks.

Empirical studies of interbank markets generally support this paper's conclusions, but add some interesting insights. Furfine (2001a) finds no evidence that interbank activity (in the federal funds market) declined following the 1998 events (Russian default, near-collapse of LTCM). Instead, he shows that, apart from the days surrounding the LTCM rescue, volumes in the second half of 1998 were higher than in the first half. He attributes this to increased dealer activity. At the same time, he finds some evidence of reduced borrowing by the most active and risky banks, which is consistent with the conclusions of Freixas and Jorge and Acharya et al.

Halsall et al (2008) find evidence that the timing of overnight loan transactions has shifted following the onset of the 2007 market turmoil. They suggest that this shift in activity was the result of lenders making their liquidity available later in the day (ie after 2pm) when they had greater certainty about their own funding needs. They further find that, after the Bank of England widened the range around the banks' target reserve bands, lending shifted back to the earlier part of the day

A third group of papers, relevant for our study, look at the importance of relationships in interbank markets. In a seminal paper, Rochet and Tirole (1996) model the monitoring role of banks in these markets. They show that banks have strong incentives to monitor each other when interbank loans are large and unsecured. But these incentives can be undermined if banks believe that large financial institutions would never be allowed to fail.

Furfine (2001b) confirms this risk monitoring role showing that US federal funds rates do indeed differentiate between banks in ways which plausibly reflect counterparty credit risk. At the same time, he finds that access to this market can rapidly dry up, partly as a result of banks' reluctance to signal to the market that they need funds, and partly because other banks wish to limit their credit risk exposure. In other words, Furfine finds evidence of credit rationing rather than an increase in the rates charged to individual banks when their condition deteriorates. Cocco et al (2003) highlight the importance of relationships in the interbank market in providing banks with insurance against liquidity risk, in the form of both unexpected shortages and surpluses of funds. Using data for the Portuguese interbank market, they find evidence of riskier borrowers relying on established relationships. Furthermore, they find that during the 1998 crisis, when overall liquidity fell, borrowers relied more than usual on banks with which they had an existing lending relationship.

Building on these insights, the present paper uses methods developed in statistical mechanics and network science to examine whether the network characteristics of the UK unsecured overnight market changed during the recent market turmoil. Statistical mechanics of networks (developed in the physics community and used to study complex networks such as the World Wide Web) has recently been applied to the study of patterns of payment and loan flows in various systems. This methodology is particularly suited for our purpose as it provides a series of summary statistics that reveal the complexities of borrowing and lending relationships at the individual bank level.

Specifically, we ask whether overall activity in the interbank market has changed since August 2007, either showing a decline related to credit rationing or an increase reflecting greater market trading activity. This is our first hypothesis. To test our second hypothesis – whether the nature of interbank flows has changed – we examine whether relationships between banks have changed and whether there is evidence of re-grouping, as some lending relationships are scaled down or others are strengthened. As explained earlier in this Section, the former could be the result of credit rationing. The latter could happen if banks chose to rely more on well-established relationships. Alternatively, banks may have decided to diversify their counterparty exposure by setting up lending relationships with more counterparties.

The remainder of the paper is organised as follows. After providing an overview of the sterling money market (Section 10.2) and the methodology (Section 10.3), we take a first look at the data in Section 10.4 and propose some conjectures about changes in the overnight market in Section 10.5. Section 10.6 tests these conjectures against network measures, and Section 10.7 assesses the robustness of our analysis. Section 10.8 concludes.

10.2 Overview of the market turmoil

Conditions in global money markets have been unusually stressed since summer 2007. Market liquidity has fallen sharply, particularly at maturities beyond one month, and spreads over policy rates have widened. Many banks have therefore found it difficult to access longer-term funding on acceptable terms. In the United Kingdom too, term money markets have seen a fall in liquidity and term spreads remain wide compared to pre-August 2007 levels. At shorter dates, however, market activity has been less impaired and rates have stayed closer to policy rates.[2]

During this period central banks have continued to provide liquidity via the normal channels (open market operations, reserves management, standing facilities) to keep short-term market rates close to policy rates. In addition, central banks have responded to the continued strains in money markets by introducing a range of extra-ordinary measures, ranging from auctions offering longer-term funding to widening collateral lists and broadening the range of counterparties (CGFS, 2008).

The UK money market framework introduced in May 2006 allowed banks to choose their own reserve targets. Each bank's average reserve balance over the maintenance period (which last four or five weeks between monthly Monetary Policy Committee meetings) had to be within a certain band around this target in order to be remunerated at Bank rate. If the balance fell short, the bank would be forced to borrow the shortfall from the Bank of England at a penalty rate. If the balance was over the target, the bank would earn no interest on the excess.

In response to the strains in sterling money markets, the Bank of England widened the range around banks' reserve targets in September 2007, thereby giving banks greater flexibility in managing their reserve accounts. In addition, banks' reserves targets in aggregate have risen since August 2007.[3] In a further reflection of the demand for short-term liquidity, UK banks have increased their own liquidity buffers in the form of high-quality collateral.

The Bank of England's operational framework also allows banks to borrow or lend on an overnight basis using the standing facilities. Initially, use of these overnight standing facilities was published on

[2] See Bank of England (2008a), chart 1.5; Bank of England (2008b), chart 23; and Bank of England (2008c), chart 2.
[3] See Bank of England (2008c), chart 30.

the following day. During summer 2007, stigma became attached to the borrowing facilities and banks ceased to use them. In October 2008, the Bank of England announced reforms to the facilities in order to reduce the stigma attached to their use.

The broad picture for 2007 and 2008 is therefore one of increased demand for short-term money (partly to replace longer-term funding and partly to create buffers), and a reduction of longer-term supply. In the remainder of this paper, we examine how these general market developments have manifested themselves in interbank relationships in the sterling overnight market.

10.3 Methodology and data

10.3.1 Network topology literature

The study of networks has been applied to a wide range of fields, such as social interactions, epidemiology and the world-wide web. A large amount of work from the physics community has focused on the structure of complex networks – that is, those displaying features that are neither regular nor purely random.[4] Recently, economists have started using these methods to analyse the patterns in payment and loan flows, and assess the stability of these networks.

Boss et al (2004) and Inaoka et al (2004) were among the first to use topology in empirical studies of interbank markets, examining the Austrian and Japanese banking and payment systems respectively. These papers confirm that topology measures are suitable to describe financial networks in general and their resilience to shocks in particular.

Soramäki et al (2007) examine the network topology of interbank payments across the Fedwire Funds Service in the US. They find that participation in the payment system fell following the attacks of 11th September 2001, both in terms of number of active banks and number of transactions. There is evidence of less coordination between banks, most likely as a result of both the operational problems faced by some participants and the responses by others to the resulting liquidity problems. However, once the operational problems were over, activity rose to above-average levels as banks settled their backlogs of payments.

---

[4] See eg Dorogovtsev and Mendes (2003), Albert and Barabási (2002) and Newman (2003).

Becher, Millard and Soramäki (2008) examine interbank payments across CHAPS Sterling. They investigate the impact on the network of an operational outage of a major settlement bank, and find both that the network topology did not look significantly different during the outage day and that non-stricken banks were able to manage their liquidity effectively so that payment flows between them were much as normal.

Papers by Pröpper et al (2008) and Lublóy (2006) use payment data from the Dutch and Hungarian payment systems, respectively. Pröpper et al find that payment activity has been considerably higher since the market turmoil of 2007 began, but network properties have been relatively unaffected. Lublóy looks at the permanence of linkages over time and finds that, although there are relatively few pairs of banks which lend to each other every single day, those that do account for the majority of payment orders by value.

Topology measures are also used in a number of recent papers which look at the overnight money market. Iori et al (2008) study the Italian money market and find that the network has changed over time. Here, banks have increased the number of counterparties they borrow from, while decreasing the number they lend to. They further document that a few large banks borrow from a large number of small counterparties. Bech and Atalay (2007) and Bonde and Bech (2008) examine overnight money markets in the US and Denmark, respectively.

10.3.2 The data

Our paper uses a set of well-established network measures. These are briefly explained in next Section. To carry out the topology analysis, we use the data on payments in the large-value payment system CHAPS Sterling available to the Bank of England in its role as operator of the underlying RTGS processor. From the raw data, it is difficult to distinguish cash payments from loan payments (either advancement of principal or repayment of principal plus interest). Following Halsall et al (2008), we use a variation of the algorithm developed by Furfine (1999) which identifies pairs of payments on consecutive days that could be interpreted as overnight loan advances and repayments. Appendix 1 discusses the construction of the data set in detail.

Our sample period runs from 18 May 2006 to 16 December 2008. To test our hypotheses, we consider a pre-crisis phase (18 May 2006 –

8 August 2007) and several crisis periods as the turmoil evolves. We identify the following key dates during the period of turbulence:

- 9 August 2007: the commonly accepted date for the start of the liquidity crisis in the UK.[5]
- 4 October 2007: the Bank of England pre-announces an increase in the bands around the reserve target at which reserves are remunerated at Bank rate.[6]
- 15 September 2008: the default of Lehman Brothers, heralding the worst period of the crisis so far.
- 8 October 2008: the UK government announces a recapitalisation scheme for UK banks.

This yields five phases for our analysis:

- Phase 0: from 18 May 2006 to 8 August 2007
- Phase 1: from 9 August 2007 to 3 October 2007
- Phase 2: from 4 October 2007 to 12 September 2008
- Phase 3: from 15 September 2008 to 7 October 2008
- Phase 4: from 8 October 2008 to 16 December 2008

As outlined in Section 10.1, we wish to test two hypotheses: i) whether overall activity in the interbank market has declined since phase 0; and ii) to what extent relationships in the interbank market have changed.

10.3.3 Interbank loan markets as networks

In the next Section, we will define our various network measures and explain how to interpret them in the context of loan markets. Before doing so, we introduce the basic network definitions. More detail is given in Appendix 2.

We model the series of unsecured loan payments between banks as an evolving network. Each bank is represented by a node, and a loan

---

[5] See, for example, Borio (2008).

[6] The bands had been widened in September 2007 too, but this was announced after the maintenance period began on 6 September 2007. Therefore banks' reserve management behaviour during the period may have changed. The maintenance period beginning 4 October 2007 was the first where banks were aware beforehand of a widening of the bands.

advance between two banks as a directed link between nodes from the lending bank to the borrowing bank.

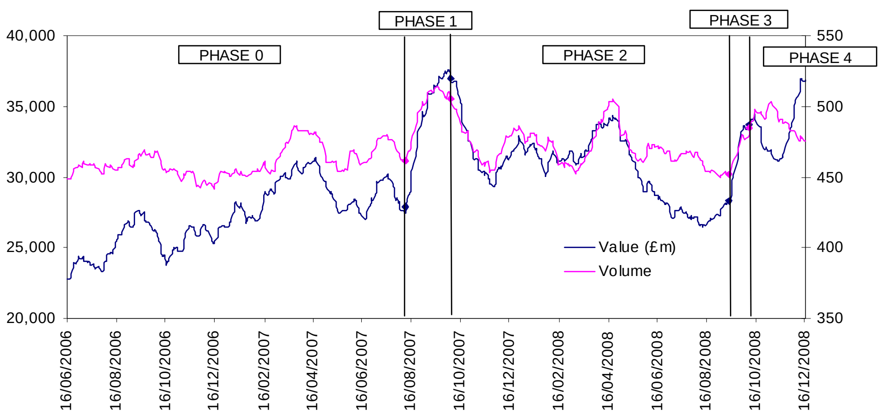
Turning to the loan markets we are interested in, we model the series of bilateral unsecured loan payments between banks as an evolving network. Each bank is symbolised by a node, and a payment between two banks by a directed link between nodes. The directed link originates from the lending bank and points to the borrowing bank. We take a business day as our unit of time, and do not look at intraday networks. This is because a bank's reserve management – and thus lending and borrowing – behaviour is driven to a large extent by its daily target.

10.4 A first look at the data

10.4.1 Values and volumes

Figure 10.1 and Table 10.1 provide a picture of the unsecured overnight sterling market. A new settlement member joined CHAPS in October 2007 and another ceased direct membership in September 2008, so we use an average daily figure.

Figure 10.1 **Total daily advances in overnight unsecured sterling – 21-day rolling average (value £m on left-hand axis; volume on right-hand axis)**



The data show that average daily values of loan advances were higher during the entire crisis period, increasing from £27.0bn in phase 0 to £33.9bn in phase 4. Despite the presence of a thirteenth bank during parts of phases 2 and 3, average daily lending per bank was higher as well. The volume of loan contracts increased too during the crisis period, but not to the same extent. During phase 2 the number of contracts per bank per day was actually slightly lower than during phase 0.

Table 10.1 **Summary statistics**

		Phase					
		All	0	1	2	3	4
	Days in phase	656	311	39	239	17	50
	Average nodes	12.37	12	12	12.99	12.29	12
Value	Daily average £m	29,522	27,015	36,699	30,333	34,659	33,897
	Per node £m	2,387	2,251	3,058	2,335	2,819	2,825
	t-statistic			13.8	3.2	6.0	11.6
	p-value			***0.0%	***0.1%	***0.0%	***0.0%
Volume	Daily average	469	460	509	469	487	487
	Per node	38	38	42	36	40	41
	t-statistic			10.8	-11.4	1.7	5.9
	p-value			***0.0%	***0.0%	10.5%	***0.0%

The t-statistics in Table 10.1 are obtained from Welch's t-test on the hypothesis that the daily mean value/volume (per node) in the phase is significantly different from the mean in phase 0.[7] We can see that in all phases the daily value per node is significantly higher than in phase 0, while volume is significantly higher in phases 1 and 4 (and significantly lower in phase 2).

Summarising the results so far, we have seen that overall overnight loan activity increased slightly post August 2007 (our first hypothesis). Of course, our data allows us only to comment on the unsecured overnight markets and not loan activity in general. In what follows, we examine to what extent relationships between counterparties have been affected by the crisis (our second hypothesis). This is done by looking at four different network

[7] Welch's t-test uses the null hypothesis of equality of means. It requires that the underlying observations are normally distributed: for the longer phases we can appeal to the central limit theorem but we should be wary of drawing strong conclusions about phase 3. In all our tables, one asterisk denotes significance at the 10% level, two for the 5% level, and three for the 1% level. Throughout this paper, we say a result is 'significant' if the null hypothesis is rejected at the 5% significance level.

measures: connectivity, reciprocity, clustering and persistence. We first define the measures and then present our results.

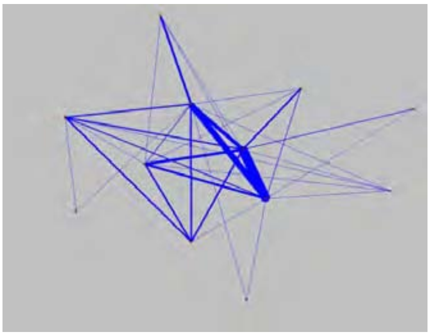
10.4.2 Network graphics

In this section we present the network graphically, using the software Pajek.[8] Obviously the structure of the network changes over time so here we display six pictures for each of the key dates (plus the start and end of the period) described in section 10.3.2. In these pictures, a link is represented by a black line between two nodes. The arrow shows the direction of the link. The thickness of the line represents the value of loans between the nodes. Where a link exists in both directions, the line is coloured blue. Note that the order of the nodes has been permuted between pictures to preserve the anonymity of individual banks.

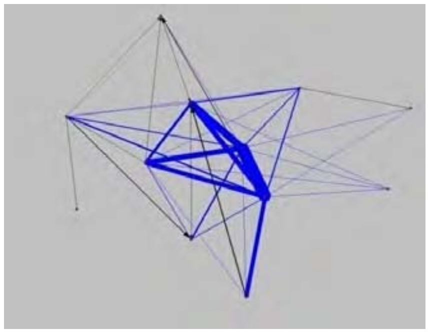
[8] For more information regarding Pajek, see http://pajek.imfm.si/doku.php?id=pajek.

Network graphics on each of the key dates

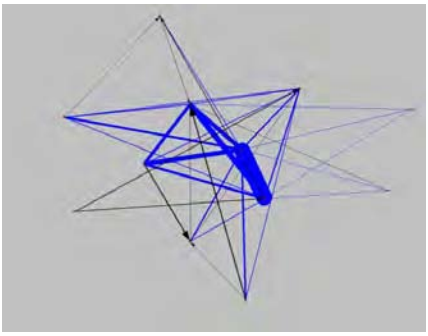
(1) 18 May 2006



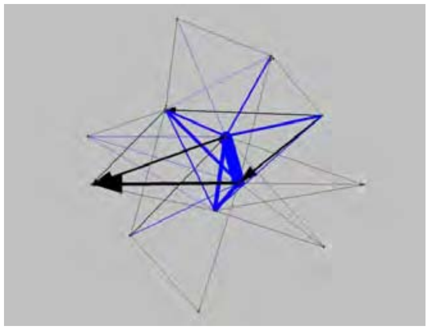
(2) 9 August 2007



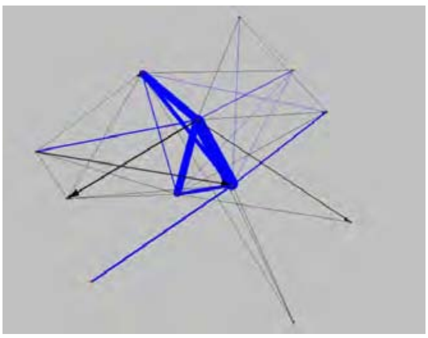
(3) 4 October 2007



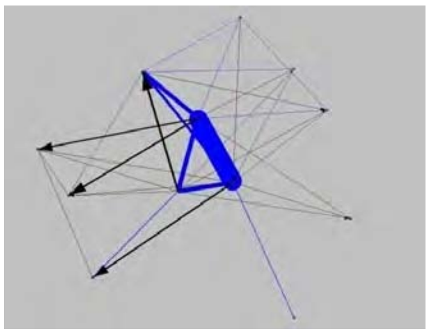
(4) 15 September 2008



(5) 8 October 2008



(6) 16 December 2008



Key

Reciprocated link

Loan (arrow points to borrower)

Bank

At first sight, the interbank network has not changed greatly. On all six days, a significant number of relationships are in place, indicating that most banks are active in the market and lending to one another. The network does not become noticeably less connected during the crisis. Furthermore, on all six days, a small number of banks dominate overnight activity, though we cannot say whether the identity of these 'core' banks remains unchanged. It appears that the proportion of connections involving the core increases as the crisis intensifies. We also observe that bilateral relationships become less reciprocal during the crisis, as there are fewer blue lines and more black. In other words, for each pair of banks, it is more likely that there is a link in one direction only. In the remainder of the paper, we use more rigorous measures to see whether these observations are an accurate description of network activity during the entire sample period.

10.5 Questions

10.5.1 Do core banks become more important?

The network graphs suggest the existence of a small group of banks which are dominant in the overnight market. For any given node, there is a higher probability of a link (either in or out) to a core node than there is to a non-core node. In other words, given any node i, any core node j and any non-core node k, we have $P(i \to j) > P(i \to k)$ and $P(j \to i) > P(k \to i)$.[9]

As explained in the introduction, theory suggests that when banks become more concerned about liquidity risk across the market, they may attempt to reduce risk exposure by relying more on established relationships. That suggests examining the data for two possible changes: (i) whether the banks increase their reliance on existing relationships as the crisis unfolds and (ii) whether the core banks benefit relatively more from this concentration. We would expect to observe both effects in our data during the crisis phases 1 and 2. Specifically, we would expect a greater proportion of links to have a core node as one or both counterparties.

We would also expect the core to become relatively less important in phases 3 and 4. There are two reasons for this. The first is that the market-wide turmoil in September and October 2008 revealed that

[9] Here $P(a \to b)$ denotes the probability that a link from node a to node b is formed. Nodes i, j and k are distinct.

even banks regarded as reliable could face problems.[10] Therefore banks may have started to rely on the core a little less, and diversified their counterparties. Once the government's recapitalisation plan was announced at the beginning of phase 4, general counterparty risk may have been adjudged to have decreased across the market.[11] This again made the core slightly less special than before. But we still expect the core to be more important than in phase 0.

10.5.2 Is there an asymmetry between lender and borrower behaviour?

Counterparty risk is a concern for both borrowers and lenders in the overnight unsecured market. If a lender develops liquidity problems, it may choose to cut down on the amount of funding rolled. But a borrower would need to have very severe liquidity problems before failing to repay its overnight loan. For a borrower, failure to repay may trigger a credit event, but a lender is free to decide not to roll a loan (unless of course it is a committed line). With this argument in mind, we examine the data to see whether money market relationship became less reciprocal and more asymmetric during the crisis. In addition, we want to see whether banks looking for a lender became more likely to choose a counterparty from the core than those looking for a borrower.

As access to term markets becomes more difficult during the crisis, borrowers may have to rely more on the overnight market, and would need to monitor the risk of their lender choosing not to roll over funding. In contrast, we would expect the risk affecting lenders to increase by a much smaller margin. Hence, it seems likely that during phases 1 to 3 borrowers would increasingly rely on the core, relative to lenders. With term markets resuming in phase 4 to a limited extent, we would expect this effect to be less pronounced.[12]

---

[10] Bank of England (2008a), p. 17–19.
[11] Bank of England (2008a), p. 29–30.
[12] Bank of England (2008a), p. 34.

10.5.3 Have the widened reserve bands had an impact?

One reason for banks to participate in the overnight market is to manage their reserve account balances close to the target, as described in Section 10.2. Although the target only binds on the last day of the maintenance period, banks actively manage their end of day balances throughout the maintenance period. If a bank is long or short near the end of the day, it may choose to lend or borrow overnight rather than miss its target. See Appendix 3.2 for changes to the bands during the period.

Since this reserve management activity behaviour is largely discretionary, we might expect it to be strongly affected by the heightened counterparty concern during the crisis. But the bands remained relatively tight at ±1% until the end of phase 1, so banks did not have much room to manoeuvre.

However, the Bank of England widened the bands in September 2007, and the first day of phase 2 (4 October 2007) is the start of next the maintenance period. Thus we might expect banks to respond much more strongly to the crisis in phases 2 and 3 than in phase 1. In phase 4 the bands start to contract again so, along with the effect of the recapitalisation plan, we might expect reaction to the crisis to be more muted in phase 4. But there still is a crisis – and bands remain wider than in phase 0 – so we should still see some impact.

In conclusion, we expect to see:

- Evidence of a core throughout the entire period;
- More dependence on the core during the crisis, and especially in phases 2–4;
- Greatest dependence on the core in phase 2;
- Borrowers in particular cutting back on non-core relationships in phases 2–4.

10.6 Analysis of network measures

In this section, we discuss our results, looking at measures of connectivity, reciprocity, clustering and persistence. We focus on network-wide measures rather than studying individual nodes, since the phases identified are defined by market-wide and not bank-specific events. For algebraic definitions of all these measures, see Appendix 2.

10.6.1 Connectivity

Definition

First, we define degree as the number of links in the network. This is a simple measure, but it has the disadvantage of not adjusting for changes to the number of nodes in the network. As the number of nodes in our network varies over time, we need a measure which takes this into account.

The connectivity of a node is the proportion of potential links that exist. Thus, for a network with n nodes, connectivity is equal to degree divided by n(n-1). Connectivity can be thought of as the probability that any given link is formed.

Results

Figure 10.2 below shows how the connectivity of the network changes over the period. The black line shows the backward-looking 21-day rolling average, while the blue diamonds denote the first day of a new phase. There appears to be a very large drop in the early part of phase 2, after which connectivity remains at a lower level before a considerable rise starting around the beginning of phase 3. Since the start of phase 4 it has been stable, at around the same level as in phase 0.

Figure 10.2 **Connectivity of the network over the period**

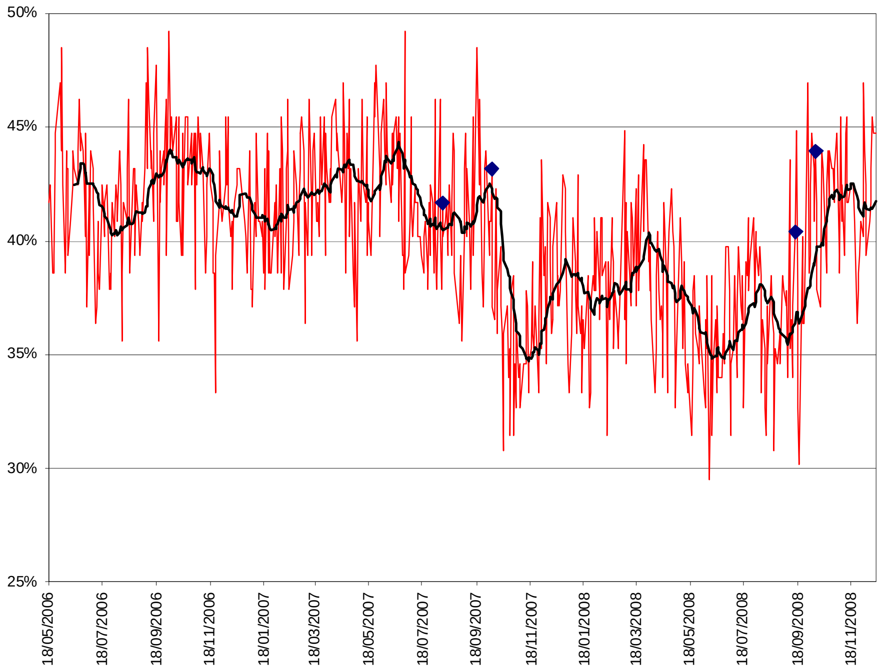


Table 10.2 below shows mean connectivity during each phase in the period, and tests for differences from the mean value in phase 0. We see that although connectivity does not decline significantly in the initial stage of the crisis (phase 1), it is considerably lower during the period between the widening of reserve target bands (start of phase 2) and the announcement of the UK government's measures (end of phase 3).

Table 10.2 **Mean connectivity during each phase**

	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4
Mean	42.10%	41.51%	37.12%	39.63%	42.05%
Variance	0.07%	0.09%	0.09%	0.18%	0.06%
t-statistic		-1.2	-20.3	-2.3	-0.1
p-value		25.0%	***0.0%	**3.2%	89.3%

Connectivity is significantly lower during much of the crisis period. This is consistent with our conjecture that banks become more wary of counterparty risk and are less likely to form relationships with non-core banks. As explained in 10.5.3, the effect is less pronounced during phase 1, when banks are more constrained in their reserves management. Connectivity returns to pre-crisis levels after the UK government announces its bank recapitalisation plans.

10.6.2 Reciprocity

Definition

Reciprocity is defined only for directed networks, and is the probability that a link between two nodes exists, given that the link in the opposite direction between the same two nodes exists.

Reciprocity can be thought of as the strength of bilateral relationships. If reciprocity is high, then banks tend to use the same counterparties for both lending and borrowing. Banks form links with the core more often than they do with non-core banks. Since the core is a relatively small group of nodes, this means that banks' borrowing and lending is likely to be correlated. When a bank chooses a lender, it is likely they will choose a member of the core, and it is likely that the core member will lend to that bank too. Thus we might expect reciprocity to be higher than connectivity, which is the probability of linking to any other bank, core or not. Moreover, we might expect it to increase relative to connectivity during the crisis period.

In fact, reciprocity does not rise during the crisis period, even though we predicted greater use of the core and thus more correlation. But this may reflect the overall fall in the probability of any relationship after October 2007 (see section 10.6.1), rather than reduced willingness to enter into borrowing and lending relationships with the same counterparties. To separate the two effects and ascertain whether reciprocity has risen relative to connectivity, we calculate a new measure called normalised reciprocity. This is calculated as reciprocity divided by connectivity.[13]

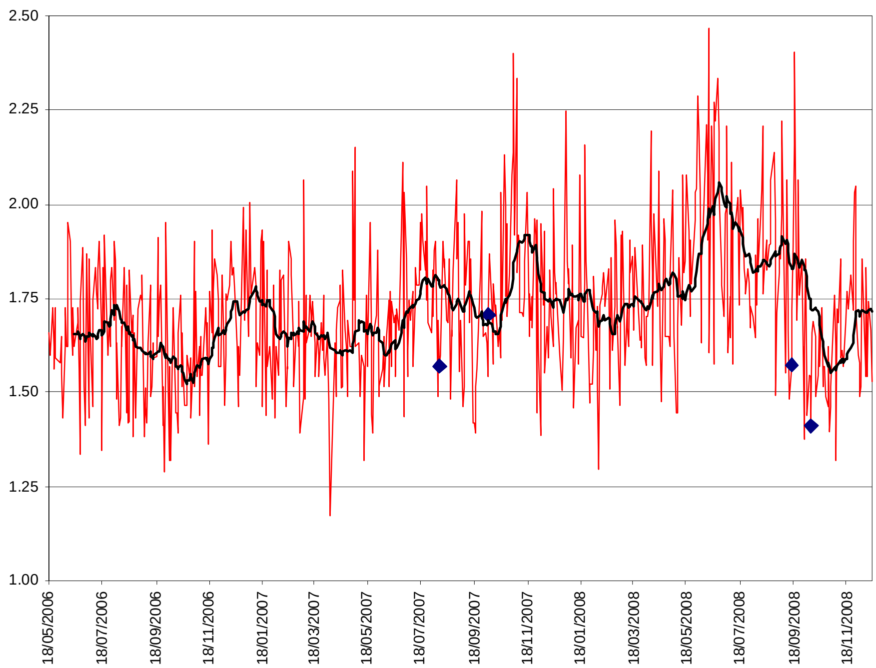
[13] Suppose we have a 'prior' of a random network – that is, one where the probability of each link existing is identical (probability p, say) and independent of the existence of other links. Then both connectivity and reciprocity have expected value p. Thus, if we observe a decrease in reciprocity, how can we distinguish between the case where p has declined, and the case where reciprocated links have become less likely (ie the network is no longer random)? An obvious solution is to 'normalise' reciprocity by dividing by connectivity. The expected value of this normalised measure would be constant and equal to 1 in a random network, and so we could tell whether the likelihood of reciprocated links has changed relative to all links.

Results

Figure 10.3 below shows how this normalised reciprocity measure has changed over the period. The Figure shows a slight increase in the second half of phase 2, ending shortly before the Lehman default (start of phase 3). Phase 3 sees a sudden drop, followed by a slight rise in the second half of phase 4.

Since normalised reciprocity is greater than 1 in all five phases, we can surmise that reciprocity is always greater than connectivity and the core plays a role even in non-crisis times. The increase in normalised reciprocity throughout the crisis means greater correlation between borrowing and lending counterparties, and could be interpreted as greater dependence on the core.

Figure 10.3 **Normalised reciprocity of the network over the period**



Normalised reciprocity in phase 4 appears closer to phase 0 levels. As suggested earlier, this could indicate that the recapitalisation plan has reduced the importance of the core, meaning that relationships with non-core banks become relatively more likely.

Table 10.3 below confirms this story, with the strongest effect seen in phase 2 (when the mean is significantly different to the phase 0 mean). In phases 1 and 3 the core banks are less important – as explained in Section 5 – and the results are not significant.

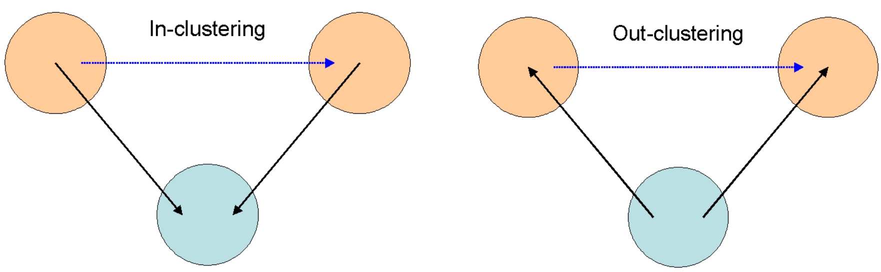
Table 10.3 **Mean normalised reciprocity during each phase**

	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4
Mean	1.66	1.71	1.81	1.75	1.64
Variance	0.03	0.03	0.04	0.07	0.02
t-statistic		1.5	9.3	1.4	-0.9
p-value		15.2%	***0.0%	18.3%	35.8%

10.6.3 Clustering

Definition

In-clustering is the probability that two nodes which both have a link to the node in question also have a link with each other (in either direction). Similarly we can define out-clustering. Since in-clustering for the whole network will not necessarily equal out-clustering, we examine them separately. In-clustering allows us to consider whether banks with a common borrower tend to lend to each other too. Similarly, out-clustering tells us the probability that banks which borrow from the same third party also lend to one another.



Consider a bank with two lenders. As the core becomes more important, it is more likely that one or both of those lenders are members of the core. There is then an increased chance of those banks forming a relationship as well. Thus in-clustering should be higher than connectivity. A similar argument applies to out-clustering. According to the argument set out in section 10.5.2, we might expect borrowers to have a greater propensity to choose counterparties from the core than lenders, and so in-clustering should be higher than out-clustering.

If we expect the core to become more important during the crisis, then in-clustering should increase relative to connectivity. A similar argument applies to out-clustering. But as borrowers are more likely to choose counterparties from the core than lenders, in-clustering should increase by more than out-clustering (relative to connectivity).

As discussed in section 10.6.2, we need to distinguish between the cases where changes in clustering are due to changes in the probability of all links being formed (ie a change in connectivity), and those where the probability of clustering has changed relative to other links. For a random network where links are formed independently with probability p, the expected values of connectivity, in- and out-clustering are all p. Thus we look at the clustering scores divided through by connectivity, which we call normalised clustering.

Results

The charts for normalised in- and out-clustering are rather dissimilar. Most noticeably, normalised in-clustering tends to be higher than out-clustering. The sharp peak in in-clustering around the start of 2008 is not seen in the out-clustering chart. The only notable common feature is an increase in the latter part of phase 4.

Figure 10.4a **Normalised in-clustering of the network over the period**

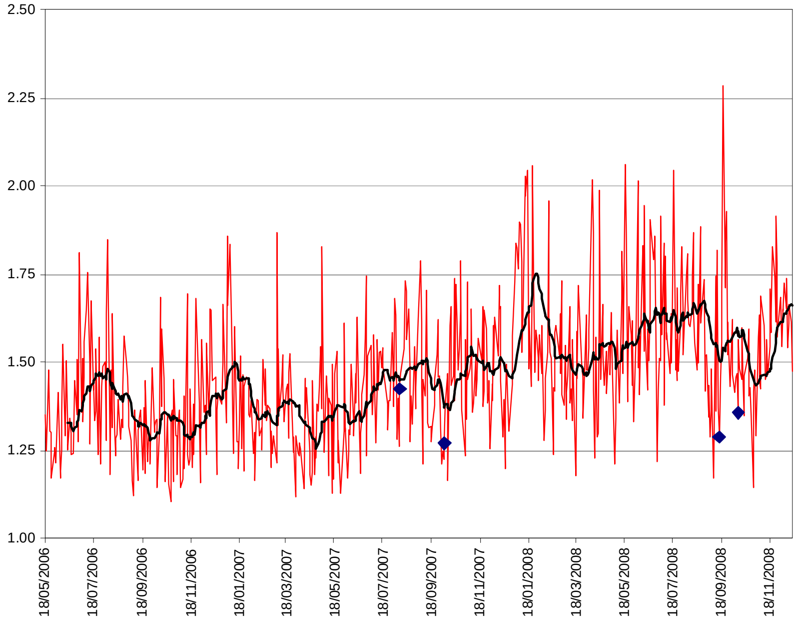
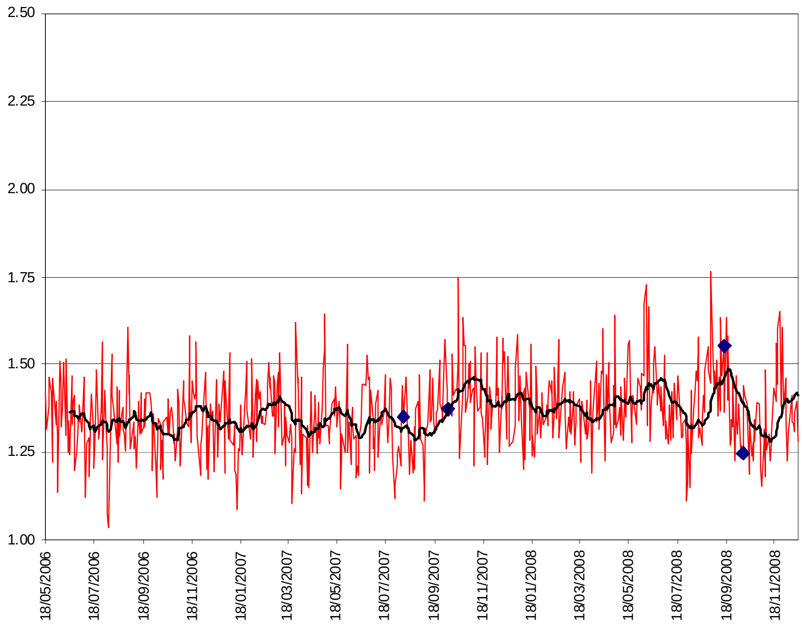


Figure 10.4b **Normalised out-clustering of the network over the period**



According to Table 10.4 normalised in-clustering is higher than phase 0 in all four of the crisis phases, while out-clustering is only significantly higher in phase 2. Thus it does appear that in-clustering is higher relative to connectivity when the core becomes more important, and the effect persists even in phase 4 when the core is not as strong.

However, normalised out-clustering is only significantly higher in phase 2, which is when we expect the core to be most important. This suggests perhaps that the probability of lending to the core does not increase by much during the crisis; the changes in connectivity and normalised reciprocity we saw earlier are driven much more by an increased probability of borrowing from the core banks rather than lending to them.

Table 10.4 **Mean normalised in- and out-clustering during each phase**

in-clustering	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4
Mean	1.37	1.45	1.56	1.60	1.55
Variance	0.02	0.02	0.04	0.06	0.02
t-statistic		2.8	12.6	3.9	8.1
p-value		***0.7%	***0.0%	***0.1%	***0.0%

out-clustering	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4
Mean	1.34	1.34	1.40	1.38	1.35
Variance	0.01	0.01	0.01	0.02	0.01
t-statistic		0.081	6.373	1.312	0.399
p-value		93.6%	***0.0%	20.7%	69.1%

Furthermore, a test for a significant difference between normalised in- and out-clustering finds that there is a very significant (at the 1% level) difference between the two. This confirms that normalised in-clustering is indeed higher than out-.clustering

10.6.4 Persistence of relations

Definition

An important part of our theory is that banks wish to choose reliable lenders as they are dependent on the rolling of overnight funding. It may also be that, during times of heightened counterparty risk, banks rely more upon established relationships so that liquidity risk can be monitored more easily – this is the insurance factor mentioned in

Section 10.1. In this sub-section we measure the propensity to roll overnight funding.

We define persistence as the probability that, given a loan is extended on day t, another loan is extended from the same lender to the same borrower on day t+1. This is particularly important in the overnight market, since we might expect that a significant proportion of overnight loans are rolled over every day. Rolling an overnight loan would involve the same lender and borrower on consecutive days. Thus persistence may be a guide to increased use of overnight funds instead of term lending in the market.

Of course, just because we observe loans between the same two parties on consecutive days it does not mean we have a rolled loan. There could be other reasons for high persistence, such as some banks tending to lend more than borrow or vice versa, or seasonal patterns leading to correlation between consecutive days. But we argue that an increase in rolling activity across the market would result in an increase in persistence.

On any given day, we argue that banks are likely to form relationships with counterparties in the core. Since the core consists of a small subset of banks, we would then expect relationships on consecutive days to be positively correlated. Thus persistence should be higher than connectivity – if the two were equal, this would suggest that there was no correlation between links on consecutive days. Moreover, we expect it to increase relative to connectivity as the core becomes more important.

As with reciprocity and clustering, we need to strip out the effect of changes in connectivity. Since the expected value of persistence in a random graph on day t is $p^{(t+1)}$, the probability of any link existing on day t+1, we define normalised persistence on day t as persistence on day t divided by connectivity on day t+1.

Results

Figure 10.5 shows that normalised persistence increases during phase 2 and drops again in phases 3 and 4. However, in all four crisis phases it appears to be higher than in phase 0. This provides evidence for the increased importance of the core during the crisis, and for a slight reduction in importance in phase 4.

Figure 10.5 **Normalised persistence of the network over the period**

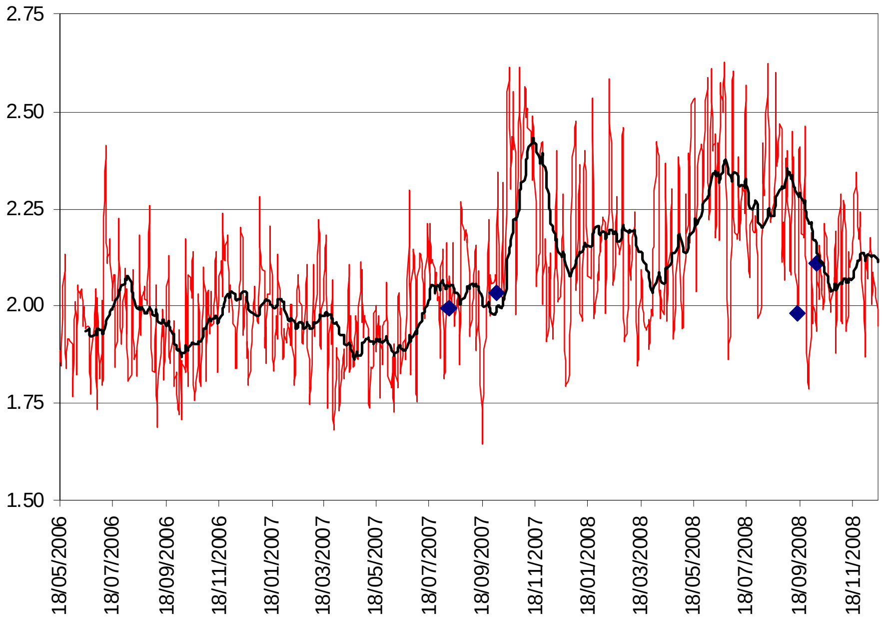


Table 10.5 shows that throughout the crisis period, normalised persistence is significantly higher than in phase 0 in every phase of the crisis. Again, the highest values occur in phases 2 and 3, when the core appears to be most important. This may also suggest that banks become more dependent on rolling overnight funding during these phases, as the term funding markets have become impaired.

Table 10.5 **Mean normalised persistence during each phase**

	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4
Mean	1.96	2.02	2.23	2.11	2.09
Variance	0.02	0.02	0.03	0.03	0.01
t-statistic		2.4	19.1	3.3	7.7
p-value		**2.2%	***0.0%	***0.4%	***0.0%

10.7 Robustness checks

So far we have argued that the network can be described as a core of a small number of banks with high connectivity, surrounded by a larger number of non-core banks. The most significant results are obtained in phases 2 and 3. These phases coincide with changes in the underlying membership of CHAPS. One bank became a direct member in October 2007 and another ceased its direct membership in September 2008. These dates are close to the start of phase 2 and the end of phase 3 respectively. It is possible that these changes have had an impact on the structure of the network.

We can attempt to control for these changes by considering a scenario where neither of the 'floating' banks was ever a direct CHAPS member, and instead all of their overnight activity was done through their former/later correspondent bank during the entire period. This gives us a network with 11 nodes. We refer to this as the 'merged network'.

We can find that connectivity in the merged network is higher than the original score most of the time. The difference is particularly obvious during the period when there were 13 banks in the network (8 October 2007 – 19 September 2008). There is no longer a significant drop at the start of phase 2 or rise at the start of phase 3.

However, there are limitations to this analysis. We do not know that banks would have kept their lending and borrowing behaviour the same if the floating banks had been second-tier members throughout the entire sample period. For example, lending or borrowing from a customer bank makes no difference to the settlement bank's reserve account balance, so its actions may well have been different when the customer was a direct member of the network.

In summary, although there is evidence that changes in CHAPS membership may have contributed to our results, we have no conclusive evidence that they alone explain the observed changes in the network since we cannot say for certain what the counterfactual situation would have been.

10.8 Concluding remarks

To summarise our analysis of the overnight market, we find no evidence of an overall reduction in overnight loan activity – in fact loan flows are up a little – but some banks may have changed their behaviour. First, our network results indicate that there exists a core of a small number of banks which account for a large portion of overnight activity. When concerns about counterparty risk increase, banks in the network prefer to borrow from or (to a lesser extent) lend to the core rather than non-core banks. However, we also observe that it was not until the reserve target bands were widened in September 2007 that banks were able to adjust their liquidity management to reduce the number of counterparties.

The turmoil in autumn 2008 raised widespread concerns about counterparty risk, reducing reliance on the core and encouraging banks to diversify their range of counterparties. And the recapitalisation plan in October 2008 reduced concerns about all banks, making it less necessary to rely on the core. Thus it is in the period October 2007 – August 2008 that we see the core at its most important.

Second, we observe a reduction in the number of relationships, indicating that the network has become less connected and less reciprocal. None of the differences are large, but they are statistically significant. To explain these changes, we suggest that during the crisis period, banks have become more concerned about counterparty liquidity risk. Many borrowers use the overnight market to obtain term lending by rolling overnight loans to the next day. Thus borrowers are keen to obtain their funds from a counterparty that is likely to agree to roll over the loan the next day. That in turn implies that they will be concerned about changes in their counterparties' liquidity risk: if the lender suffers a liquidity shock then they might be less likely to agree to roll over the loan.

Counterparty liquidity risk is a concern for lenders too. If a borrower cannot find the funds to repay a loan, the lender will either have to roll over the loan or consider the borrower to be in default. But being in default sends a very negative signal to the market, and a borrower is likely to seek to avoid this at all costs. Thus the borrower will repay the loan if it is able to. In contrast, a lender suffering liquidity constraints can decide not to roll over a loan without (direct) adverse effects. Hence, we argue that increased concerns about counterparty risk affect choice of lender more than choice of borrower.

We do not attempt to measure whether the impact of market events was greater or less than the impact of policy events. This question could be important when attempting to gauge the efficacy of central bank actions. For example, increased access to central bank liquidity as a result of policy changes during the crisis may have crowded out private provision, affecting the relationships that private banks have with each other.

In conclusion, our network analysis of the overnight money market indicates that the structure of relationships between banks changed as the crisis unfolded. But the analysis also suggests that the observed changes were small, and that underlying trading activity was unaffected. More work is needed to understand how activity in the overnight unsecured market was affected by changes in the term markets and in the secured markets, and what conclusions we can draw about the resilience of liquidity in the money markets and the need for infrastructure changes. We leave this for future research.

References

Acharya, V V – Gromb, D – Yorulmazer, T (2008) **Imperfect competition in the interbank market for liquidity as a rationale for central banking.** LBS Working paper.

Albert, R – Barabási, A L (2002) **Statistical mechanics of complex networks.** Review of Modern Physics, Vol. 74, 47–97.

Allen, F – Carletti, E – Gale, D (2008) **Interbank market liquidity and central bank intervention.** Conference proceedings Liquidity: Pricing and Risk Management, June 2008.

Ashcraft, A – McAndrews, J – Skeie, D (2008) **Precautionary reserves and the interbank market.** Mimeo Federal Reserve Bank of New York.

Bank of England (2008a) **Financial Stability Report.** October 2008.

Bank of England (2008b) **Quarterly Bulletin.** Vol. 48 No. 3.

Bank of England (2008c) **Quarterly Bulletin.** Vol. 48 No. 4.

Bech, M L – Atalay, E (2007) **The topology of the federal funds market.** Mimeo Federal Reserve Bank of New York.

Becher, C – Millard, S – Soramäki, K (2008) **The network topology of CHAPS sterling.** Bank of England 2008.

Bonde, K – Bech, M L (2008) **The topology of the Danish money market.** Mimeo Federal Reserve Bank of New York.

Borio, C (2008) **The financial turmoil of 2007 – ?: a preliminary assessment and some policy considerations.** BIS working paper 251, March 2008.

Boss, M – Elsinger, H – Summer, M – Thurner, S (2004) **The network topology of the interbank market.** Quantitative Finance, Vol. 4 (6), 677–684.

Brunnermeier, M (2008) **Deciphering the 2007-8 liquidity and credit crunch.** Forthcoming Journal of Economic Perspectives.

Caballero, R J – Krishnamurthy, A (2007) **Collective risk management in a flight to quality episode.** NBER Working paper No. 12896.

Cocco, J F – Gomes, F J – Martins, N C (2003) **Lending relationships in the interbank market.** LBS IFA Working paper No. 384.

Committee on the Global Financial System (2008) **Central bank operations in response to the financial turmoil.** CGFS Publications No. 31.

Dorogovtsev, S N – Mendes, J F F (2003) **Evolution of networks.** Oxford University Press.

Flannery, M J (1996) **Financial crises, payment system problems, and discount window lending.** Journal of money, credit and banking, Vol. 28 (4), 804–824.

Freixas, X – Jorge, J (2008) **The role of interbank markets in monetary policy: A model with rationing.** UPF Working paper No. 1027.

Freixas, X – Parigi, B M – Rochet, J C (2000) **Systemic risk, interbank relations, and liquidity provision by central banks.** Journal of money, credit and banking, Vol. 32, No. 3, part 2, 611–638.

Furfine, C (1999) **The microstructure of the federal funds market.** Financial markets, institutions and instruments, Vol. 8, No. 5, 24–44.

Furfine, C (2001a) **The interbank market during a crisis.** BIS Working Paper No. 99.

Furfine, C (2001b) **Banks as monitors of other banks: Evidence from the overnight federal funds market.** Journal of business, 2001, Vol. 74, No. 1, 33–57.

Halsall, C – Jackson, J – Merrouche, O (2008) **The impact of reserves averaging on banks' management of payment shocks.** Mimeo, Bank of England.

Inaoka, H – Ninomiya, T – Shimizu, T – Takayasu, H – Taniguchi, K (2004) **Fractal network derived from banking transaction – an analysis of network structures formed by financial institutions.** Bank of Japan Working Paper No. 04-E-04.

Iori, G – de Masi, G – Precup, O V – Gabbi, G – Caldarelli, G (2008) **A network analysis of the Italian overnight money market.** Journal of Economic Dynamics and Control, Vol. 32 (1), 259–278.

Lublóy, A (2006) **Topology of the Hungarian large-value transfer system.** Magyar Nemzeti Bank Occasional Paper No. 57.

Millard, S – Polenghi, M (2004) **The relationship between the overnight interbank unsecured loan market and the CHAPS Sterling system.** Bank of England Quarterly Bulletin, Spring 2004.

Newman, M E J (2003) **The structure and function of complex networks.** SIAM Review, Vol. 45, 167–256.

Pröpper, M – van Lelyveld, I – Heijmans, R (2008) **Towards a network description of interbank payment flows.** De Nederlandsche Bank Working Paper No. 177.

Rochet, J C – Tirole, J (1996) **Interbank Lending and Systemic Risk.** Journal of Money, Credit, and Banking, Vol. 28, 733–762.

Soramäki, K – Bech, M L – Arnold, J – Glass, R J – Beyeler, W E (2007) **The topology of interbank payment flows.** Physica A, Vol. 379, 317–333.

Appendix 1

Finding overnight loans from payment data

The Furfine method for finding overnight loans employed in this paper can be summarised as follows:

1. Generally loan principal amounts are in fairly round numbers. On day t, find all payments in round numbers and label them as possible overnight loan advances. For overnight loans, we define a ‘round number’ as being of value £1m or above and divisible by £100,000. Thus we only consider loans of value £1m or more.

2. On day t+1 label all payments of value £1m or above and not a ‘round number’ as possible overnight loan repayments. For each of these, calculate the implied principal amount by rounding down to the nearest £100,000, or round down to the nearest £1m if the repayment amount is greater than £250m.[14]

3. Match possible advances on day t with implied principal amounts from repayments on day t+1. For each potential matched pair, check:

   a. The advance has not already been matched with another repayment, and vice versa;
   b. The payer of the advance matches the payee of the repayment, and vice versa;
   c. The implied interest rate falls within ±2% of the Bank of England rate;
   d. The implied interest rate is plausible, meaning that is an exact number of basis points or half-points. To allow for rounding errors, we accept interest rates within 0.01 basis points of an exact number.

[14] There may appear to be a danger here of making an error if the repayment amount is slightly more than a multiple of £100,000 or £1m. But if a loan of size ≤ £250m and the interest is ≥ £100k, then the implied annualised interest rate is over 14%, which is unreasonable considering prevailing interbank rates in the UK during the period studied. And if a loan commands interest of ≥ £1m, then even at a rate of 8% (very high compared to 3-month Libor over the period), the principal amount would be over £4.5bn. Payments of this magnitude are very rare, and there have been none on consecutive days between the same pairs of banks over the period we are examining.

This algorithm is relatively accurate for overnight loans. It is unlikely to be 100% accurate since it could happen that opposing payments on consecutive days look like a loan advance and repayment without being part of loan. Also there is a problem of ambiguity if there are payments on three or more consecutive days that could be advances and repayments (in this case the algorithm assumes an advance on the first day and repayment on the second). But the likelihood of these coincidences is low enough to justify use of the method. We also lose any overnight loans that are less than £1m in value, but again this should not have a major impact. Halsall et al (2008) conduct robustness checks on this data set and confirm that the data are representative of the sterling unsecured overnight market.

But there is a further complication in that the CHAPS database includes only payments made between the 13 clearing banks which are CHAPS members, and therefore excludes loan payments between direct and indirect participants and loan payments between two customers of the same settlement bank, which are settled across that bank's books rather than in CHAPS. Data on these are not available.

Appendix 2

Algebraic definitions of topology measures

Nodes: Given n nodes, we label them {1,2,…,n}. The topology measures are invariant under permutations of this labelling.

Links: If there is a link from node i to node j, then $l_{ij} = 1$. If there is no such link, then $l_{ij} = 0$. As no node can have a link to itself, $l_{ii} = 0$ for all i.

For the entire network, we compute the following

Connectivity $\chi = \dfrac{\sum_i \sum_j l_{ij}}{n(n-1)}$

Reciprocity $r = \dfrac{\sum_i \sum_j l_{ij} l_{ji}}{\sum_i \sum_j l_{ij}}$

In-clustering $c^{in} = \dfrac{\sum_i \sum_j \sum_k l_{ji} l_{ki} l_{jk}}{\sum_i \sum_j \sum_k l_{ji} l_{ki}}$

Out-clustering $c^{out} = \dfrac{\sum_i \sum_j \sum_k l_{ij} l_{ik} l_{jk}}{\sum_i \sum_j \sum_k l_{ij} l_{ik}}$

Define $l_{ij}^{(t)}$ as the value of l_{ij} on day t.

Then persistence $\pi^t = \dfrac{\sum_i \sum_j l_{ij}^{(t)} l_{ij}^{(t+1)}}{\sum_i \sum_j l_{ij}^{(t)}}$

Normalised measures are defined as the relevant measure divided by χ.

Appendix 3

Appendix 3.1

Major events during the 2007-08 UK market turmoil

Date	Event
July 07	First real signs of a crisis. Bear Stearns announces that two subprime hedge funds have rapidly declined in value.
9 Aug 07	Generally accepted start of crisis. ECB and Federal Reserve inject around £45bn of funds into the financial markets.
20 Aug 07	First use of standing facilities during the crisis; clear signs that borrowing on facilities carries stigma.
13 Sep 07	Bank of England announces a widening of reserve bands for the current maintenance period.
14 Sep 07	Northern Rock granted liquidity support facility from Bank of England.
8 Oct 07	UBS joins CHAPS as a direct member.
17 Mar 08	JP Morgan Chase agrees to buy Bear Stearns.
21 Apr 08	Bank of England launches Special Liquidity Scheme, which allows commercial banks to borrow Treasury bills in exchange for less liquid collateral.
15 Sep 08	Bankruptcy of Lehman Brothers.
17 Sep 08	Disclosure of merger talks between HBOS and Lloyds TSB.
22 Sep 08	ABN Amro ceases its direct membership of CHAPS.
29 Sep 08	Bradford & Bingley, a UK mortgage bank, part-nationalised.
8 Oct 08	Announcement of UK bank recapitalisation plan.
20 Oct 08	Reform of standing facilities.

Appendix 3.2

Range around reserve targets within which reserves are remunerated at Bank rate

Announcement	Band	Effective	Announcement	Band	Effective
18 May 06	±1%	18 May 06	18 Sep 08	±40%	4 Sep 08
13 Sep 07	±37.5%	6 Sep 07	1 Oct 08	±60%	4 Sep 08
20 Sep 07	±60%	6 Sep 07	6 Oct 08	±40%	9 Oct 08
2 Oct 07	±30%	4 Oct 07	3 Nov 08	±20%	6 Nov 08
7 Jul 08	±20%	10 Jul 08	1 Dec 08	±10%	4 Dec 08

Chapter 11

An agent-based model of payment systems*

Marco Galbiati – Kimmo Soramäki

* This article first appeared as Bank of England Working Paper n:o 352. The publisher gratefully acknowledges the permission to reprint this article.

11 An agent-based model of payment systems

Abstract

This paper lays out and simulates 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. The paper studies 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 and ii) insights on the efficiency of alternative system configurations.

Summary

A large share of all economic transactions is ultimately settled via money transfers between banks, taking place on 'large-value payment systems' (LVPSs). In 2006, the annual value of interbank payments made in the European system TARGET totalled €533 trillion (about $670 trillion), amounting to more than 50 times the value of the corresponding countries' gross domestic products. The sheer size of these transactions, and their importance for the functioning of the economy, explains why policymakers are interested in LVPSs, and in the behaviour of their participants.

In the past, most payment systems worked on a deferred, net settlement basis. During a business day the banks would exchange promises of payments, deferring the actual transfer of funds to the end of the day, when only net positions were settled. The advantage of this arrangement was that only net debtors had to actually provide funds, and only in a quantity sufficient to cover their *net* position. Because net positions are typically small (compared to gross payments), the system as a whole would require little liquidity to function. Today instead, most LVPSs work on a gross settlement basis: there is no netting, and a payment obligation is legally discharged only when the corresponding full amount is transferred across accounts held at a central bank. This apparent backward step, strongly encouraged by monetary authorities worldwide, was motivated by credit risk concerns. Suppose indeed that, in a net system, at the end of the day a bank is unable to make good its final position. Its creditors may face losses too large to be sustained, so their payments too might have to be cancelled, creating a domino effect with significant consequences for financial stability. Gross settlement eliminates this risk but requires more liquidity, as the benefits (not only the risks) of netting are foregone. These arguments suggest that the provision of liquidity is an essential issue to modern payment systems.

Real-time gross systems are more ‘liquidity hungry’ than deferred net systems. However, they allow for liquidity ‘recycling’: when a bank receives a payment, it can use the received funds to make other payments of its own. To make an analogy, in a football game the ball can be passed between the players many times; similarly, a same unit of liquidity can be used to settle many payments. Consider however what happens if the ball is expensive to buy – maybe no one would like to pay for it in the first place. Unfortunately the analogy carries on to payment systems, where liquidity (the ball) bears a cost for commercial banks. This is an interest cost (typically charged by the central bank) or an opportunity cost (when liquidity is obtained against a pledge of collateral). So, even though just a little liquidity could generate a large volume of exchanges, it is unclear who should provide it. Banks are thus faced with a dilemma: to act as liquidity providers by acquiring costly funds, or to wait for liquidity to arrive from other banks. In the first case a bank does not depend on its partners, and it can promptly execute payments. In the second case, a bank benefits from a free source of liquidity, but is exposed to the risk of delaying payments while waiting for funds to arrive.

This paper develops a dynamic model of liquidity provision in a payment system, where banks face a choice between: a) the costs of borrowing liquidity, and b) the cost of delaying payments. In more detail, the model is a sequence of days. At the beginning of each day, every bank chooses how much liquidity to borrow from external sources. This liquidity is then used to execute payment orders which arrive throughout the day in a random, exogenous fashion (these orders can be interpreted as being commissioned by a bank’s external clients, or by some area of the bank, different from the treasury). As long as the bank has sufficient funds, payments are executed as soon as they are received; when instead a bank’s liquidity balance reaches zero, payments are queued until incoming payments provide the bank with new funds. Finally, at the end of the day banks receive profits, which depend on the liquidity borrowed, and on the delays suffered in executing payment orders. Day after day, banks adapt their liquidity choices following a particular learning process. As a consequence, the banks' behaviour eventually stabilises, and the banks end up providing an equilibrium amount of liquidity.

The system's equilibrium level depends on the model's parameters. By changing these, we look at the amount of liquidity absorbed by the system in a variety of scenarios, drawing conclusions on the efficiency of the system. We find that, for a wide range of costs, efficiency could be enhanced if banks were to commit more liquidity than they do in equilibrium. This might constitute a rationale for imposing measures that encourage liquidity provision (for example, throughput guidelines). From a different perspective, systems with fewer participants are found to be more liquidity-efficient than larger ones, due to the emergence of 'liquidity pooling' effects, as described by previous studies. These results are found by varying the *size* of the system but not its *structure*: it is outside the scope of this work to look at how liquidity choices are affected by changes in the extent of 'tiering' of a payment system (that is, we do not fully investigate the case of banks 'moving out' of the system, and making their payments through other system participants).

11.1 Introduction

Virtually all economic activity is facilitated by transfers of claims by financial institutions. In turn, these claim transfers generate payments between banks whenever they are not settled across the books of a (perhaps third) institution. These payments are settled in interbank payment systems. In 2006, the annual value of interbank payments made in the European system TARGET totalled €533 trillion (about $670 trillion). In the corresponding US system Fedwire, the amount was $572 trillion, while the UK system CHAPS processed transactions for a value of £59 trillion (about $109 trillion). In perspective, these transfers amounted to 24 to 40 times the value of the respective countries' GDPs. The sheer size of the transfers, and their pivotal role in the functioning of financial markets and the implementation of monetary policy, make payment systems a central issue for policymakers 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); also, a payment can be submitted only if the paying bank has enough funds to deliver the full amount in central bank money (gross settlement). Because no netting takes place, RTGS modality imposes high liquidity demands on the banks, making RTGS systems vulnerable to *liquidity risk*, ie to the risk that liquidity-short banks are unable to send their own payments. This may create delays and possibly cause gridlocks in the system (see eg Bech and Soramäki, 2002). Hence, liquidity is one of the central issues 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. To our knowledge, this is the first paper that explores this question using an 'agent-based' approach, ie combining elements of game theory and numerical simulations.

The amount and the distribution of liquidity in a payment system is the result of a complex interaction between the system's participants. Indeed, during the day, each bank has to make a stream of payments, that can only be partly predicted. To cover the liquidity needs generated by these payments, banks typically rely on two sources: a) reserve balances or credit acquired from the central bank and b) funds received from other settlement banks during the course of the day. The first source can be seen as providing *external* (to the system) *liquidity*, while the second is a source of *internal liquidity*. In normal conditions a bank can draw freely on external liquidity. This however has a cost, which gives incentives to economise on its use.[1] 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 is also costly.[2] As a consequence, a bank has to optimally decide how much external liquidity to acquire, trying to forecast when and how much internal liquidity it will receive, trading off external liquidity costs against (expected) delay costs. The fact that banks i) delay some payments, and yet ii) do not wait till the very end of the day to make all their payments, shows that this trade-off indeed exists.

Two main difficulties emerge when studying the behaviour of banks in a payment system. First, when modelled in sufficient detail, liquidity flows in RTGS systems follow complex dynamics, making

[1] The costs of acquiring liquidity are opportunity costs (returns that the bank would obtain if it could employ this liquidity differently), and interest costs (costs from borrowing the liquidity itself).

[2] Delays usually carry two types of cost. First, formal agreements often penalise late delivery; if a delay extends over the end of the due day, penalties may apply. Second, delays may entail reputational costs, which are difficult to quantify but potentially large.

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 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 spillover effects on the rest of the system, so no system participant can solve its optimal liquidity demand problem in isolation. As strategic interactions are widespread, banks interact in a fully fledged 'game', jointly determining the performance of the system.

This paper studies this liquidity game, putting particular effort into modelling liquidity flows. 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 approach and instead to use simulations. In particular, we use numerical methods to compute a crucial element of the game, the pay-off function, or a relationship between i) a bank's own external liquidity, ii) the external liquidity of other banks, and iii) the resulting settlement delays and costs.

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 over time, using an adaptive process whereby actions are chosen on the basis of past experience. We then simulate the resulting dynamics and we look at the limit, or equilibrium, behaviour. This depends on the specific form of the adaptive rule, so we choose the 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. A convergence point of our dynamics will be a Nash equilibrium of the liquidity game.

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 (2005). These papers model various 'liquidity management games' with a few agents and a small number of periods (respectively, two and three). While these models improve our understanding of the incentives in payment systems, the actual pay-off functions may be too simple to describe costs in real payment systems accurately. As we said, in RTGS systems liquidity can circulate many times and between many banks, generating dynamics that cannot be captured by these simple, but analytically tractable, models.

Recently, a growing literature has used simulation techniques to investigate efficiency and risk issues payment systems (see eg James and Willison (2004) and the volumes edited by Leinonen (2005, 2007)). 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 either been 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.

Recognising the strengths and disadvantages of these two approaches, the present paper tries to build a bridge between them, combining the strength of each of them. Of course, we have to leave something behind: the realism of historical data (which may however be inappropriate to study counterfactual scenarios), and the sharpness of analytical results.

The paper is organised as follows: Section 11.2 provides a formal description of the model, describes some properties of the cost function, and illustrates the *tatônnement* process towards equilibrium. Section 11.3 presents the results of the experiments and Section 11.4 concludes.

11.2 Description of the model

The model is a stylised representation of a day in RTGS, where the banks (players) engage in the following game.

11.2.1 Banks and liquidity choices

At the beginning of the day, each of N banks (denoted by $i = 1...N$) chooses its reserves, say $l_i(0)$, to be used in the course of the settlement day.[3] To simplify, we assume that these reserves, the external liquidity, can only be acquired once, at the beginning of the day. Once reserves have been (simultaneously) chosen, the settlement day begins: banks start receiving payment orders, and execute them

[3] In the simulations, we assume that $l_i(0)$ is an integer between 0 and some large L.

using available liquidity. In game-theoretic terms, $l_i(0)$ is bank i's action and the vector $l = (l_1(0), l_2(0)...l_N(0))$ is an action profile. The next subsection illustrates how payments are received and executed, generating the outcome of the game.

11.2.2 Payments and delays

The outcome of the day-game is determined as follows. The day is modelled as a continuous time interval [0,T]. Payment orders arrive according to a Poisson process with parameter $\lambda = 1$, so the system as a whole receives, on average, T orders per day. The payor and the payee of these payment orders are determined by (uniform) random draws: for any order, the probability that banks i and $j \neq i$ are respectively the payor and the payee is $(1/N)(1/(N-1))$. Equivalently, each single bank receives payment orders according to a Poisson process with parameter $\lambda = 1/N$, and the payee of each such order is determined by a random (uniform) draw. These orders can be seen as generated *outside* the bank, by a bank's clients, or *within* the bank, by some area which is different from the treasury department. Whatever the interpretation, payment orders are exogenous for the agent choosing $l_i(0)$.

Let us call $z_i(t)$ the number of payment orders *received* by bank i up to time t, and $x_i(t)$ the number of payment orders *executed* by i up to t. At t, bank i's queue (its backlog of outstanding orders) is therefore

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

where 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

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

where $y_i(t)$ is the amount of payments that i has received from other banks up to time t. For simplicity, we assume that every i adopts the following payment rule[4]

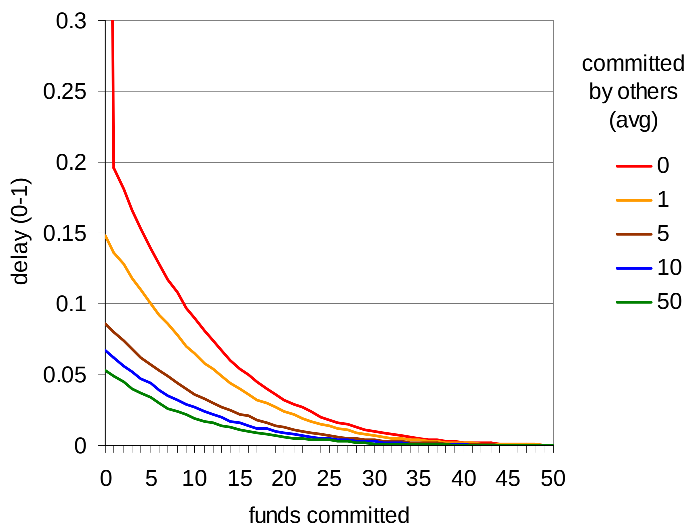
[4] Such a rule is optimal for the cost specification given in the next section: banks need to pay upfront for liquidity, so they have no incentive to delay payments if liquidity is available. Under other cost specifications (eg heterogeneous payment delay costs) this would, however, not be the case.

$$\begin{aligned}&\text{if } l_i(t) > 0, \text{ execute new and queued payments as FIFO;}\\&\text{if } l_i(t) = 0, \text{ queue new payment orders}\end{aligned} \tag{11.1}$$

Bank i's incoming payments $y_i(t)$ are just other banks' outgoing payments, so the settlement process is fully described by the above equations.

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 $l = (l_1(0),...,l_N(0))$ pins down the initial conditions of the system; from there, the exogenous arrival of payment orders mechanically generates liquidity fluxes, queues and delays. All this can be numerically simulated, to determine how delays depend on liquidity choices. For example, Figure 11.1 shows the (average) amount of delays obtained for different levels of total liquidity in the system, when $l_i(0)$ is the same for each i.[6]

Figure 11.1 **Delays as a function of total liquidity**



As system liquidity is reduced, delays increase non-linearly due to what are often referred to as 'deadweight losses' (Angelini, 1998) or

[5] Queues do not form only when $l_i(0)$ is very high. Then, $\Delta x = \Delta z$ so executed payments essentially follow a Poisson process which mirrors the arrival of payment orders.

[6] Delays are normalised such that 1 reflects a situation where all payments are delayed until the end of the day, and 0 a situation where no delays take place.

‘gridlocks’ (Bech and Soramäki, 2002). Intuitively, 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. These delay chains are more likely and more extended the lower the liquidity in the system. Thus, the total effect of liquidity reduction acts in a compounded fashion.

11.2.3 Costs

At the end of the settlement day, banks receive pay-offs that depend on the liquidity posted at the beginning of the day, and on the delays generated by the settlement algorithm illustrated in the above section. More precisely, we assume that acquiring initial liquidity $l_i(0)$ imposes a liquidity cost equal to

$$C(l_i(0)) = \lambda l_i(0), \quad \lambda > 0 \tag{11.2}$$

This is the first component of a bank’s pay-off (cost). We then suppose that a payment order received at t and executed at t’ carries a penalty equal to

$$c(t',t) = \kappa(t' - t), \quad \kappa > 0 \tag{11.3}$$

Such penalties are summed over all received payment orders of the day, to give a bank’s delay cost. A bank's total pay-off is then the sum of delay and liquidity costs.

The random arrival of payment orders generates random delays; hence, pay-offs too are a random function of the action profile l(0). As anticipated above, the analytical form of this pay-off is exceedingly complex to determine; hence, we simulate the settlement process many times for every action profile, to obtain a numerical estimate of expected costs, as a function of l(0).[7] The resulting pay-off function is plotted in Figures 11.2 and 11.3 for two levels of delays costs; ‘low’ (11.2) and ‘high’ (11.3).

[7] We assume banks are risk-neutral, ie they care about expected pay-offs.

Figure 11.2 **Costs as a function of own initial funds – low delay costs**

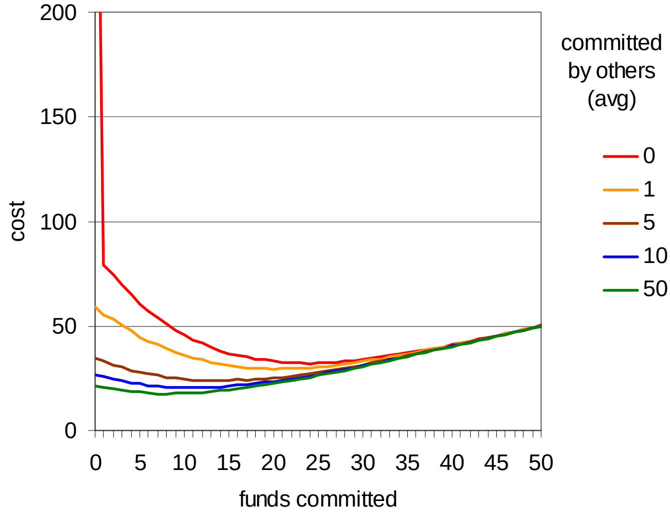
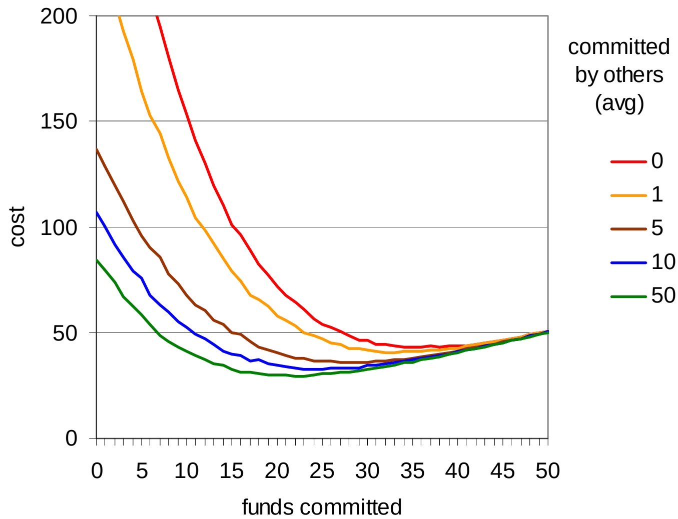


Figure 11.3 **Costs as a function of own initial funds – high delay costs**



The simulations also show an interesting fact:

Remark 1 Bank i's cost is (essentially) a function of its own action and of the sum of others' actions.

This important empirical finding, which probably depends on the fact that the day is 'long', greatly simplifies the analysis.[8] First, in certain respects it leaves us with a game with only two players: bank i playing against 'the rest of the system'. Second, it allows us to derive some analytical results, to be discussed in the next section.[9]

We find this result by comparing two sets of simulations. In the first set, we vary the total amount of liquidity, while spreading it uniformly across banks (ie we simulate the settlement day for different values of $\Sigma l_i(0)$, imposing every time $l_i(0) = (1/N)\Sigma l_i(0)\ \forall i$). In the other, we change again the total liquidity, but we distribute it randomly across banks, so $l_i(0)$ varies across banks. Comparing the total costs in the two sets, we found that the differences are small – around 2% or less. We suspect this can be explained by two facts: i) the assumption of a complete symmetric network (every bank exchanges payments to any other with similar intensity), and ii) the relatively large number of payments quickly redistributes liquidity, flushing out the initial conditions. Both assumptions are realistic in many systems, for example in the UK CHAPS system (see Soramäki et al, 2007).

11.2.4 Equilibrium

To find the equilibrium of the liquidity game, we use the so-called *fictitious play* tatônnement process (Brown, 1951). 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 the appendix; the reason to adopt this particular dynamic is twofold. First, despite its simplicity the fictitious play rule is in a sense rational and thus not too unrealistic, corresponding to Bayesian updating of beliefs about others' actions.[10] Second, fictitious play can indeed be a useful tool to compute equilibria. Indeed, when fictitious play converges to a stable

[8] We do not have a rigorous proof, but we suspect the following. When many payments are made (ie the day is 'long'), liquidity is soon spread among banks according to a stable distribution. Hence the initial distribution does not matter, only the total liquidity does.
[9] Games with this property are known as aggregation games. They have the convenient feature that a number of adjustment dynamics applied to them are 'well behaved' (see eg Mezzetti and Dindo, 2006).
[10] See eg Fudenberg and Levine (1998, page 31) for details.

action profile, this is a Nash equilibrium of the underlying game.[11] Summing up, fictitious play can be seen either as a computational device, or as a 'story' with an appealing economic meaning.

A key question is whether the game has a unique equilibrium and, if not, which equilibrium will be uncovered 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 the simulations will pick one or the other), all equilibria are characterised by exactly the same total level of liquidity $\Sigma_i l_i(0)$, which can therefore be rightly called the equilibrium liquidity. This allows us to perform comparative statics, where we change parameters of the cost function and other elements of the model.

11.3 Results

11.3.1 Liquidity demand and efficiency of the equilibrium

We start with a base case scenario with 15 banks; this number is chosen so that our system 'looks like' the UK CHAPS.[12] In all of the simulations banks interact in a complete network, ie each sends payments to every other bank in the system – another fairly realistic assumption for CHAPS.

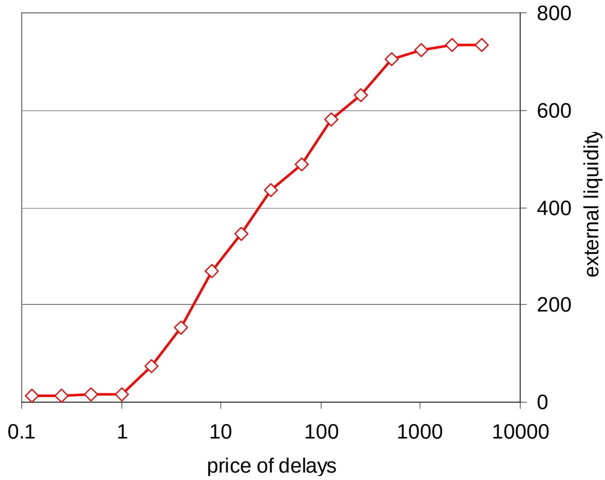
First, we obtain a 'liquidity demand function', relating the (equilibrium) amount of external liquidity $\Sigma_i l_i(0)$ to unit delay costs κ, for λ normalised to 1.[13] As expected, the amount of liquidity acquired by the banks is low for relatively inexpensive delays (Figure 11.4). When k grows, so does liquidity demand, roughly following a logarithmic pattern up to a certain point. However, as liquidity grows, delays become increasingly rare. As a consequence, decreasing returns on liquidity eventually prevail, causing liquidity demand to eventually flatten out.

[11] It is well known that fictitious play may fail to converge. However this is not the case here, as shown by the simulations. Interestingly, convergence in aggregation games was shown by Kukushkin (2004) for a dynamic similar to fictitious play.

[12] The length of the day is 3,000 'time ticks', so on average each bank makes 200 payments a day.

[13] Only λ/κ matters for the banks' decisions; hence our demand function is essentially equivalent to a 'traditional' liquidity demand, where the demand $\Sigma_i l_i(0)$ depends on the cost λ.

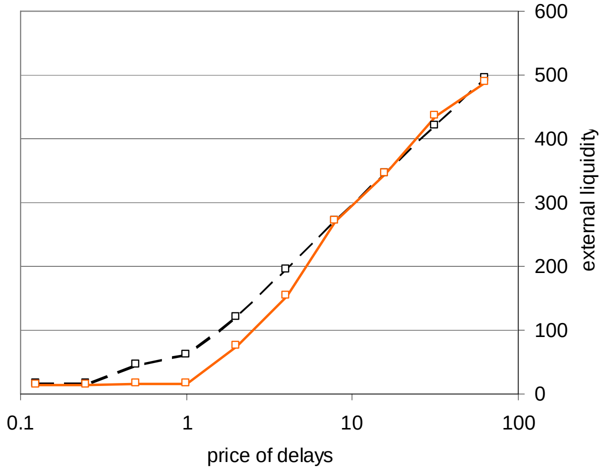
Figure 11.4 **Equilibrium external liquidity as a function of delay costs**



An important question is whether the equilibrium of the liquidity game is efficient; that is, whether the self-interested behaviour of banks can be improved upon by some co-ordinated action. To answer this question, one should ideally compare the equilibrium outcome of the game, to what would result if banks were *jointly* minimising the *total* costs of the system. To simplify computations, we search for optimal liquidity levels under the constraint $l_i(0) = l_j(0)\ \forall i,j$ – all banks are given the same amount of funds.[14] We find that the equilibrium outcome roughly coincides with the collective cost minimising choice for extreme values of k (delay costs), as shown in Figure 11.5 (the continuous line represents the liquidity minimising total cost). However, for intermediate unit delay costs, the outcome reached by independent banks is dominated by the co-ordination outcome, where more liquidity is provided as a whole.

[14] This constraint should not be binding: returns to own liquidity are decreasing, so redistribution from a liquidity-rich to a liquidity-poor bank should on average reduce total delays. Hence, an efficient allocation of liquidity should assign the same $l_i(0)$ to all banks.

Figure 11.5 **Cost-minimising common action (dashed) versus Nash-equilibrium outcome**



At the origin of such inefficiency are positive externalities in liquidity provision: external liquidity is used by all banks, but of course an individual institution only cares about private costs and benefits. Competitive banks then free-ride on others, leading to insufficient provision of external liquidity.

11.3.2 Relative efficiency of different size networks

Is a system with more participants preferable to a smaller one? This question can be considered from different points of view: from a risk / financial stability perspective,[15] or from a cost-efficiency perspective. Here, we concentrate on the second aspect. We then run experiments varying the number of banks in the model. In the first experiment, we increase the number of participants while keeping the number of payments per bank constant. In the second experiment, we increase the number of participants while keeping constant the system-wide number of payments (so per-bank payments are decreased). We measure efficiency using the netting ratio, that is the average amount of *external* liquidity required for each payment:

[15] For example, fewer participants could imply that the failure of one bank implies disruption of a larger share of payments. On the other hand, fewer participants might also mean safer participants, making it non-trivial to draw financial stability conclusions.

$$\text{netting ratio} = \frac{\text{total external liquidity}}{\text{total payments}}$$

The lower the netting ratio, the higher is the level of 'liquidity recycling' in the system.

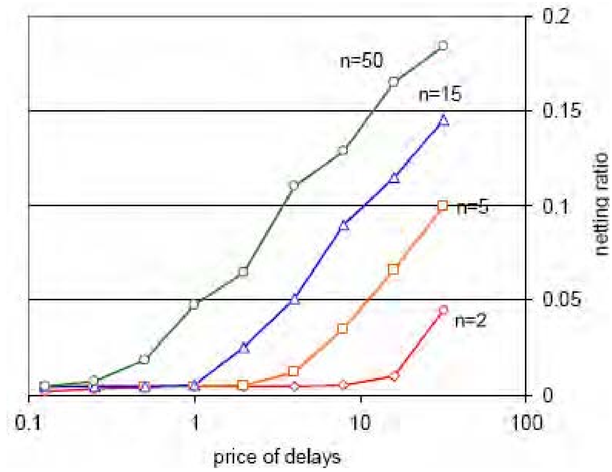
A caveat: while we change the size of the system, we maintain the assumption of a complete and symmetric payment network. This type of change (a pure 'rescaling') is convenient to analyse, but is just a simplified description of what happens in real payment systems. There, changes in the number of banks are usually accompanied by changes in the topology of the system, as some banks *de facto* merge their payment activity with others (giving rise to the so-called 'tiering'). When this happens, liquidity demand is influenced in a complex way by a number of factors, that we do not need to consider in our simplified 'rescaling' case. The interaction of liquidity demand (and costs) and tiering is outside the scope of this paper; it is instead studied in Jackson and Manning (2007).

11.3.2.1 Size effects I – constant individual bank payments

Here we vary N (the number of banks), while keeping the number of payments per bank constant – so the number of system-wide payments changes accordingly. The number of system participants has a dramatic effect on liquidity choices and efficiency. As the system size increases, liquidity demand grows while efficiency falls, and increasingly so as delays become expensive. As Figure 11.6 shows, for low delay costs the netting ratio[16] is virtually unaffected by the network size. But, at higher unit delay costs, differences are amplified and systems with fewer participants are more liquidity-efficient than systems with a higher number of participants.

[16] The average liquidity required for each payment, or the ratio (total bank external liquidity) / (total payments).

Figure 11.6 **Equilibrium external liquidity with alternative system sizes and fixed turnover per bank**



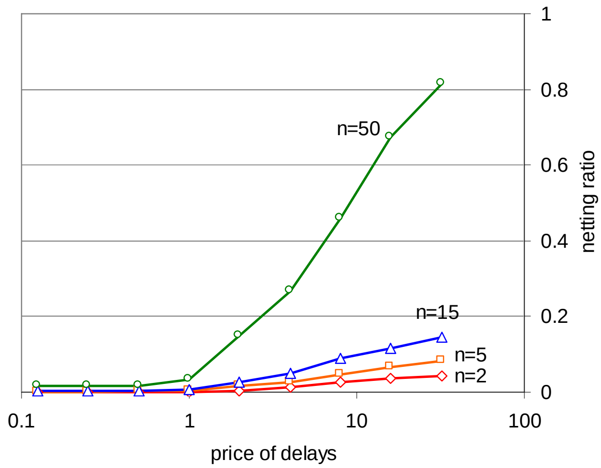
The following is an intuitive explanation of this result (as we said, liquidity flows are too difficult to be described analytically, so we can only rely on intuition to interpret the simulations). Consider a bank i and suppose N is increased from, say, 2 to 3. Because the number of payments per bank are kept constant and equally distributed over all banks, both outgoing and incoming expected payments remain constant for i in any time interval.[17] However, the *variance* of i's incoming payments increases: at each t, a bank i can now receive 0, 1 or 2 payments instead of only 0 or 1. Faced with a more unstable source of internal liquidity, the banks find it convenient to rely more on external liquidity.

11.3.2.2 Size effects II – constant total volume

In a second experiment, we keep constant the system total volume, distributing it over a varying number of banks. Note however that we keep constant the number of payments *between* banks. The results, illustrated in Figure 11.7, show a pattern similar to the previous case: systems with fewer members are seen to absorb less liquidity.

[17] If Z is the number of a bank's outgoing payments, the total outflow out of all $j \neq i$ is $(N-1)Z$. By construction, i captures a fraction $1/(N-1)$ of this flow, ie Z, which is kept constant.

Figure 11.7 **Equilibrium external liquidity with alternative system sizes and fixed total turnover**



Our (again, intuitive) explanation of this finding is as follows. A reduction of the number of banks from N to say N' can be seen as taking place in two 'steps': first, a reassignment of the payments as to involve N' banks only; second the elimination of the banks left with no payments. The second stage is neutral, as the eliminated banks are 'dummy'. Instead, the first stage brings about liquidity savings, due to the so-called liquidity 'pooling effect' (see eg Jackson and Manning, 2007). In turn, the liquidity pooling can be explained as follows: suppose the payments to/from two different banks are settled by one bank only. The volatility of the liquidity balance of this one bank increases, but by a factor less than two. Thus, a liquidity buffer of less than two times the original liquidity buffers is sufficient to settle all payments; a more precise explanation is given in the appendix.

11.4 Conclusions

In this paper we build and simulate an agent-based model of an RTGS system, paying special attention to the complex liquidity flows exchanged by the participating banks. The simulations demonstrate 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 system, and to perform various comparative statics exercises.

First, we retrieve a liquidity demand function, relating the system's liquidity to the costs faced by banks in their payment activity (liquidity versus delay costs). Then we consider the question of whether such liquidity demand, expressed by non-cooperating banks, is efficient. We find that, for a wide range of costs, efficiency (measured by the netting ratio) could be *enhanced* if banks were to commit *more* liquidity than they do in equilibrium. This might constitute a rationale for imposing measures that encourage liquidity provision (for example, throughput guidelines). From a different perspective, systems with fewer participants are found to be more liquidity-efficient than larger ones, due to the emergence of 'liquidity pooling' effects, as described by previous studies. We privileged complexity and realism, over analytical solvability. Consequently, we used a numerical, agent-based approach. Besides being useful when closed-form results are difficult to obtain, our approach is flexible and modular, allowing the present work to be extended to alternative scenarios. Further research may look at different network structures, at more elaborated liquidity management rules, at banks that differ in their costs or payment orders. Finally, our model of a 'vanilla' RTGS system could be easily extended to 'hybrid' systems like the European TARGET, which features liquidity-saving mechanisms.

References

Angelini, P (1998) **An analysis of competitive externalities in gross settlement systems.** Journal of Banking and Finance, Vol. 22, 1–18.

Bech, M L – Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory, Vol. 109 (2), 198–219.

Bech, M L – Garratt, R (2006) **Illiquidity in the interbank payment system following wide-scale disruptions.** Federal Reserve Bank of New York Staff Report No. 239.

Bech, M L – Soramäki, K (2002) **Liquidity, gridlocks and bank failures in large value payment systems.** E-money and Payment Systems Review, Central Banking Publications, London.

Beyeler, W – Bech, M – Glass, R – Soramäki, K (2007) **Congestion and cascades in payment systems.** Physica A, Vol. 384, Issue 2, 693–718.

Brown, G W (1951) **Iterative solutions of games by fictitious play.** In Koopmans, T C (ed), Activity analysis of production and allocation, New York: Wiley.

Buckle, S – Campbell, E (2003) **Settlement bank behaviour and throughput rules in an RTGS payment system with collateralised intraday credit.** Bank of England Working Paper No. 209.

Fudenberg, D – Levine, D K (1998) **The theory of learning in games.** MIT Press, Cambridge, Massachusetts.

Jackson, J – Manning, M J (2007) **Central bank intraday collateral policy and implications for tiering in RTGS payment systems.** DNB Working Paper No. 129.

James, K – Willison, M (2004) **Collateral posting decisions in CHAPS Sterling.** Bank of England Financial Stability Review, December, 99–104.

Kukushkin, N S (2004) **Best response dynamics in finite games with additive aggregation.** Games and Economic Behaviour, Vol. 48, 94–110.

Leinonen, H (2005) (ed) **Liquidity, risks and speed in payment and settlement systems – a simulation approach.** Bank of Finland Studies, E: 31.

Leinonen, H (2007) (ed) **Simulation studies of liquidity needs, risks and efficiency in payment networks.** Bank of Finland Studies, E: 39.

Mezzetti, C – Dindo, M (2006) **Better-reply dynamics and global convergence to Nash equilibrium in aggregative games.** Games and Economic Behaviour, Vol. 54, 261–292.

Soramäki, K – Bech, M L – Arnold, J – Glass, R J – Beyeler, W E (2007) **The topology of interbank payment flows.** Physica A, Vol. 379, 317–333.

Willison, M (2005) **Real-Time Gross Settlement and hybrid payments systems: a comparison.** Bank of England Working Paper No. 252.

Appendix

Fictitious play

Consider a sequence of daily games (settlement days) running from t = 0 to potentially infinity. The actions chosen on day t are a vector $l^t = \{l_1^t, l_2^t, ..., l_N^t\}$.[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 $p_i^t(.) = (p_i^t(1), p_i^t(2)...)$, where $p_i^t(x)$ is the probability that i attaches to $\sum_{j \neq i} l_j^t = x$ being played at t.
- a bank updates its belief according to the following rule:

$$p_i^t(k) = \frac{1 + \sum_{s=1...t-1} I_k(s)}{t + \Lambda}$$

  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 $\sum_{j \neq i} l_j^i = k$, and zero otherwise.[19]
- at t, bank i chooses $l_i^t = \arg\max_l \sum_{x=1}^{L} f_i(l, x) p_i^t(x)$ – where $f_i(l,x)$ is the cost incurred by i playing l, if the others play $\sum_{j \neq i} l_j^t = x$.

Equilibria in the simulations

Most of the equilibria found with the simulations have banks switching between two or more actions, depending on the evolution of their beliefs. This is due to the fact that, in the simulations, liquidity choices are discrete. For example, at the lowest delay price level banks oscillate between l = 0 and l = 1, chosen with probabilities 8.6% and 91.4%, respectively. As banks become sufficiently confident that

---

[18] Here 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.

[19] On the first day (t=0), all banks believe that each $\Sigma_{j\neq i} l_j^i$ is equally likely: $p_i^0(k) = 1/\Lambda$. Then, the more frequently a $\Sigma_{j\neq i} l_j^i$ is played, the more frequently it is 'believed' to be played again.

other banks chose $l = 1$ each, the best reply is $l = 0$. As the probability of others choosing $l = 0$ is thereby increased, banks switch back to $l = 1$. In this case, the game is a classic ‘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 bank’s 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 pure equilibrium emerges, where each bank chooses $l = 1$. At higher cost levels banks either reach a pure equilibrium, or a mixed equilibrium where they mix between a narrow range of different liquidity levels.

Uniqueness of equilibrium liquidity level

We now show that all equilibria feature the same level of aggregate liquidity. This allows us to speak about *the* equilibrium liquidity, even though the game may possess many different equilibria.

Recall that

- $f(l_i,L_{-i})$ is the expected pay-off (cost) of bank i at strategy profile (l_i,L_{-i}).

By Remark 1 (page 325), we can also consider $f(l_i,L_{-i})$ a function of two variables. So, from now on L_i is no longer a vector but a scalar, $L_i = \Sigma_{j\neq i}l_j$. We need some new notation:

- $l_i^*(l_{-i})$ is bank i's best reply to L_i.
- $\Delta_i = L_i^{'} - l_i$, to be used when $l_{-i}^{'}$ and L_i are clear from the context. Similarly, $\Delta_i^* = l_i^*(l_{-i}^{'}) - l_i^*(l_{-1})$, and $\Delta_{-i} = \sum_{j\neq i}(l_j^{'} - l_j)$ and $\Delta =_{i\in N} (l_i^{'} - l_i)$.
- $z(l_i,L_{-i})$ is the amount of delays suffered by i at strategy profile (l_i,L_{-i}), so total pay-offs are $f = \lambda l_i + \kappa z(l_i,L_{-i})$.

We can now prove our result:

Theorem 1 All equilibria feature the same total liquidity.

Proof. The argument proceeds in two steps.

Step (1) For each L_{-i} and l'_{-i}, we have $\Delta_i^* \le -\Delta_{-i}$.

That is, a bank optimally 'under-reacts' to a change in others' liquidity. To show this, first note that when we take second derivatives of $f = \lambda l_i + \kappa z(l_i, L_{-i})$ only the second term survives, so eg $\frac{\partial^2 f(l_i, l_{-i})}{\partial l_i \partial l_{-i}} = \kappa \frac{\partial^2 z(l_i, l_{-i})}{\partial l_i \partial l_{-i}}$.[20] The diagram on next page shows how the liquidity balance of a bank may evolve in time (kinked line). Delays z are measured by the area below the zero liquidity line (balances cannot become negative, so the 'depth' below the zero line represents the length of a queue). From the picture it is evident that $\frac{\partial^2 z(l_i, l_{-i})}{\partial l_i^2} < 0$, as also found in the simulations (Figure 11.1). Hence, l_i^* satisfies the first-order condition $\frac{\partial z(l_i, l_{-i})}{\partial l_i} = g(l_i, l_{-i}) = \lambda$, and the standard result applies: $\frac{dl_i^*}{dl_{-i}} = \frac{\partial g(.)}{\partial l_{-i}} \bigg/ \frac{\partial g(.)}{\partial l_i}$. Close examination of the diagram also reveals that $\frac{\partial^2 z(l_i, l_{-i})}{\partial l_i \partial l_{-i}} \le \frac{\partial^2 z(l_i, l_{-i})}{\partial l_i^2}$, so $\frac{dl_i^*}{dl_{-i}} \le -1$, which is the statement of Step (1)[21]

Step 2) If l and l' are *equilibria*, then $\Delta l = \sum l'_i - \sum l_i = 0$.

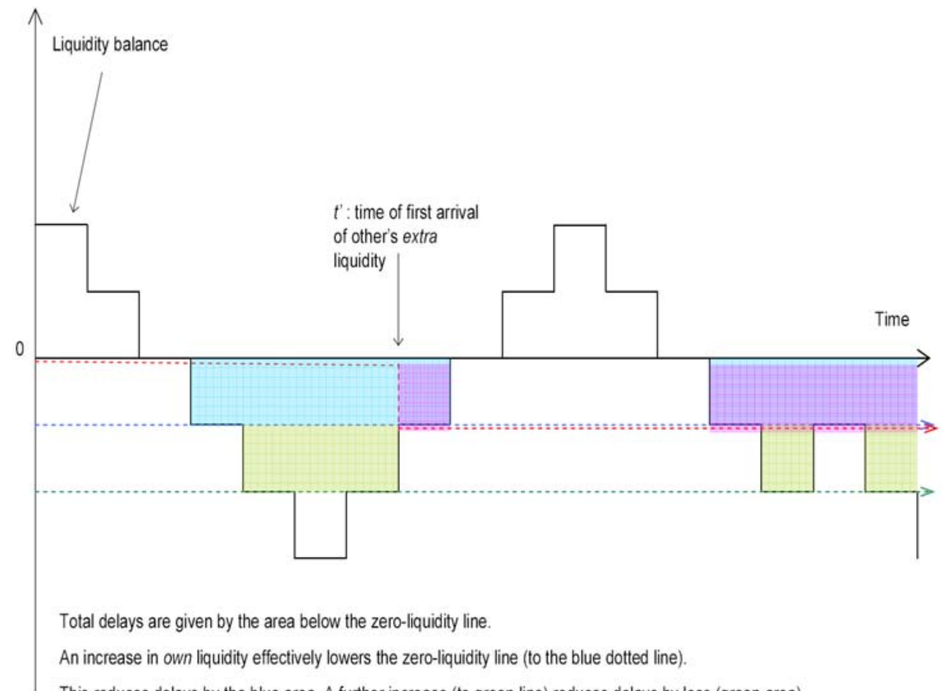
To reach a contradiction, suppose l and l' are equilibria but $\sum l'_i > \sum l_i$ ie $\Delta > 0$. If it were so, there should be a non-empty set of banks S: $\Delta_k > 0$ for all $k \in S$. By Step 1) we can write $\Delta_k = -(\Delta_{-k} + \varepsilon_k)$ (with $\varepsilon_k \le 0$), so the total change in liquidity between the two equilibria is

$$\Delta = \left[\sum_{k \in S} \Delta_k\right] + \left[\sum_{k \in N \backslash S} \Delta_k\right] = -\left[\sum_{k \in S} (\Delta_{-k} + \varepsilon_k)\right] + \left[\sum_{k \in N \backslash S} \Delta_k\right]$$

[20] Strictly speaking, we should not be using derivatives, as payments and liquidity choices are discrete. The argument in terms of difference is similar, just more cumbersome.

[21] Because $\frac{\partial^2 z(l_i . l_{-i})}{\partial l_i \partial l_{-i}} < \frac{\partial^2 z(l_i . l_{-i})}{\partial l_i^2}$ everywhere, this inequality extends to non-infinitesimal changes in L_{-i}.

The mechanics of liquidity and queues



Total delays are given by the area below the zero-liquidity line.

An increase in *own* liquidity effectively lowers the zero-liquidity line (to the blue dotted line).

This reduces delays by the blue area. A further increase (to green line) reduces delays by less (green area).

An increase in *others'* liquidity has the same effect, but only from $t'>0$. Reduction in delays is given by the pink area (smaller than blue area).

D^2z / dl_i^2 is expressed by the difference (green area) – (blue area)

$D^2z / dl_i dl_{-i}$ is expressed by the difference (green area) – (blue area), but only after t'.

Now, given a set $S = \{x_1, x_2, x_3, ...\}$ it is clear that $\Sigma_{i\in S} x_{-i} = (|S| - 1)\Sigma_{i\in S} x_i$. Similarly, if x_{-i} comes from a larger set $R \supseteq S$, then $\Sigma_{i\in S} x_{-i} = (|S| - 1)\Sigma_{i\in S} x_i + |S|\Sigma_{i\in R\setminus S} x_i$. So the above expression can be written as

$$
\begin{aligned}
\Delta &= -\left[(|S|-1)\sum_{k\in S}\Delta_k + |S|\sum_{k\in N\setminus S}\Delta_k - \overbrace{\sum_{k\in S}\varepsilon_k}^{\varepsilon}\right] + \left[\sum_{k\in N\setminus S}\Delta_k\right] \\
&= -\left[|S|\left(\sum_{k\in S}\Delta_k + \sum_{k\in N\setminus S}\Delta_k\right) - \sum_{k\in S}\Delta_k - \varepsilon\right] + \left[\sum_{k\in N\setminus S}\Delta_k\right] \\
&= (1-|S|)\Delta l + \varepsilon \\
&\Rightarrow \Delta = \frac{\varepsilon}{|S|}
\end{aligned}
$$

But $\varepsilon \le 0$, so this contradicts $\Delta > 0$.

System size and pooling effect

In the main text we said that when payments are distributed over more banks, the liquidity needs of the system increase. This is due to the liquidity pooling effect, that we now illustrate for the (simpler) case where liquidity is abundant, so queues do not form.

When liquidity is abundant, a bank's net liquidity balance is a random walk: over a time interval Δt, on average, $p\Delta t$ payments are made (pushing 'down' the liquidity balance), and $p\Delta t$ payments are received (pushing 'up' the balance). Hence, the average balance change is zero, with a standard deviation $\sigma = \sqrt{p\Delta t}$. Suppose the number of participants N is increased to $N' = Nx$ (with $x > 1$), but turnover is kept constant. Payments are now distributed over more banks, so their arrival rate is reduced from p to p/x. As a consequence, the variance in a bank's balance is reduced to $\sigma' = \sqrt{p / x\Delta t} > \sigma\sqrt{1/x}$.

Suppose now that a bank's optimal liquidity l_i is proportional to its balance variance (say $l_i = z\sigma$, which is exactly the case if a bank chooses l_i as to cover z standard deviations from the average balance). Then, the fall in variance (factor $\sqrt{1/x}$) is not enough to offset the increase in system's size (factor x), so the larger system absorbs more liquidity: $N'z\sigma' = (Nx)z\sigma\sqrt{1/x} > Nz\sigma$.

Bank of Finland Publications

Scientific monographs

Series E (ISSN 1238-1691, print) (ISSN 1456-5951, online)
From year 2009 new ISSN numbers (ISSN-L 1798-1077, print) (ISSN 1798-1085, online)

(Series E replaces the Bank of Finland's research publications series B, C and D.)

E:1 Jukka Vesala **Testing for Competition in Banking: Behavioral Evidence from Finland**. 1995. 206 p. ISBN 951-686-447-3.

E:2 Juha Tarkka **Approaches to Deposit Pricing: A Study in the Determination of Deposit Interest and Bank Service Charges.** 1995. 166 p. ISBN 951-686-457-0.

E:3 Timo Tyrväinen **Wage Determination, Taxes, and Employment: Evidence from Finland.** 1995. 212 p. ISBN 951-686-459-7.

E:4 Sinimaaria Ranki **Realignment Expectations in the ERM: Causes and Measurement.** 1996. 164 p. ISBN 951-686-507-0.

E:5 Juhana Hukkinen **Kilpailukyky, ulkomaankaupan rakenne ja taloudellinen kasvu** (Competitiveness, structure of foreign trade and economic growth). 1996. 134 p. ISBN 951-686-512-7.

E:6 Eelis Hein **Deposit Insurance: Pricing and Incentives.** 1996. 120 p. ISBN 951-686-517-8.

E:7 Vesa Vihriälä **Banks and the Finnish Credit Cycle 1986–1995.** 1997. 200 p. ISBN 951-686-537-2.

E:8 Anne Brunila **Fiscal Policy and Private Consumption-Saving Decisions: European Evidence.** 1997. 147 p. ISBN 951-686-558-5. (Published also as A-131, Helsinki School of Economics and Business Administration, ISBN 951-791-225-0, ISSN 1237-556X)

E:9 Sinimaaria Ranki **Exchange Rates in European Monetary Integration**. 1998. 221 p. ISBN 951-686-564-X.

E:10 Kimmo Virolainen **Tax Incentives and Corporate Borrowing: Evidence from Finnish Company Panel Data.** 1998. 151 p. ISBN 951-686-573-9. (Published also as A-137, Helsinki School of Economics and Business Administration, ISBN 951-791-290-0, ISSN 1237-556X)

E:11 Monica Ahlstedt **Analysis of Financial Risks in a GARCH Framework**. 1998. 181 p. ISBN 951-686-575-5.

E:12 Olli Castrén **Fiscal-Monetary Policy Coordination and Central Bank Independence**. 1998. 153 p. ISBN 951-686-580-1.

E:13 Antti Ripatti **Demand for Money in Inflation-Targeting Monetary Policy**. 1998. 136 p. ISBN 951-686-581-X.

E:14 Risto Koponen – Kimmo Soramäki **Intraday Liquidity Needs in a Modern Interbank Payment System. A Simulation Approach**. 1998. 135 p. ISBN 951-686-601-8.

E:15 Liisa Halme **Pankkisääntely ja valvonta. Oikeuspoliittinen tutkimus säästöpankkien riskinotosta** (Banking regulation and supervision: A legal policy study of risk taking by savings banks). 1999. XLIV + 560 p. ISBN 951-686-606-9, print; ISBN 951-686-607-7, online.

E:16 Juha Kasanen **Ilmoitusvelvollisten osakeomistus ja -kaupat Helsingin Pörssissä** (Corporate insiders shareholdings and trading on the HEX Helsinki Exchanges). 1999. 146 p. ISBN 951-686-630-1, print; ISBN 951-686-631-X, online.

E:17 Mikko Spolander **Measuring Exchange Market Pressure and Central Bank Intervention.** 1999. 118 p. ISBN 951-686-645-X, print; ISBN 951-686-646-8, online.

E:18 Karlo Kauko **The Microeconomics of Innovation: Oligopoly Theoretic Analyses with Applications to Banking and Patenting**. 2000. 193 p. ISBN 951-686-651-4, print; ISBN 951-686-652-2, online. (Published also as A-166, Helsinki School of Economics and Business Administration, ISBN 951-791-442-3, ISSN 1237-556X)

E:19 Juha Kilponen **The Political Economy of Monetary Policy and Wage Bargaining. Theory and Econometric Evidence**. 2000. 180 p. ISBN 951-686-665-4, print; ISBN 951-686-666-2, online.

E:20 Jukka Vesala **Technological Transformation and Retail Banking Competition: Implications and Measurement.** 2000. 211 p. ISBN 951-686-695-6, print; ISBN 951-686-696-4, online. (Published also as A-184, Helsinki School of Economics and Business Administration, ISBN 951-791-518-7, ISSN 1237-556X)

E:21 Jian-Guang Shen **Models of Currency Crises with Banking Sector and Imperfectly Competitive Labor Markets.** 2001. 159 p. ISBN 951-686-711-1, print; ISBN 951-686-712-X, online.

E:22 Kari Takala **Studies in Time Series Analysis of Consumption, Asset Prices and Forecasting**. 2001. 300 p. ISBN 951-686-759-6, print; ISBN 951-686-760-X, online.

E:23 Mika Kortelainen **Edge: a model of the euro area with applications to monetary policy.** 2002. 166 p. ISBN 952-462-001-4, print; ISBN 952-462-002-2, online. (Published also as A-204, Helsinki School of Economics and Business Administration, ISBN 951-791-715-5, ISSN 1237-556X)

E:24 Jukka Topi **Effects of moral hazard and monitoring on monetary policy transmission.** 2003. 148 p. ISBN 952-462-031-6, print; ISBN 952-462-032-4, online.

E:25 Hanna Freystätter **Price setting behavior in an open economy and the determination of Finnish foreign trade prices.** 2003. 84 p. ISBN 952-462-045-6, print; ISBN 952-462-046-4, online.

E:26 Tuomas Välimäki **Central bank tenders: three essays on money market liquidity auctions**. 2003. 232 p. ISBN 952-462-051-0, print; ISBN 952-462-052-9, online. (Published also as A-218, Helsinki School of Economics, Acta Universitatis Oeconomicae Helsingiensis, ISBN 951-791-762-7, ISSN 1237-556X)

E:27 Heikki Hella **On robust ESACF identification of mixed ARIMA models**. 2003. 159 p. ISBN 952-462-112-6, print; ISBN 952-462-113-4, online.

E:28 Heiko Schmiedel **Performance of international securities markets**. 2004. 275 p. ISBN 952-462-132-0, print; ISBN 952-462-133-9, online.

E:29 Tuomas Komulainen **Essays on financial crises in emerging markets**. 2004. 173 p. ISBN 952-462-140-1, print; ISBN 952-462-141-X, online.

E:30 Jukka Vauhkonen **Essays on financial contracting**. 2004. 134 p. ISBN 952-462-172-X, print; ISBN 952-462-173-8, online.

E:31 Harry Leinonen (ed.) **Liquidity, risks and speed in payment and settlement systems – a simulation approach**. 2005. Compilation. 350 p. ISBN 952-462-194-0, print; ISBN 952-462-195-9, online.

E:32 Maritta Paloviita **The role of expectations in euro area inflation dynamics**. 2005. 88 p. ISBN 952-462-208-4, print; ISBN 952-462-209-2, online.

E:33 Jukka Railavo **Essays on macroeconomic effects of fiscal policy rules**. 2005. 150 p. ISBN 952-462-249-1, print; ISBN 952-462-250-5, online.

E:34 Aaron Mehrotra **Essays on Empirical Macroeconomics**. 2006. 243 p. ISBN 952-462-290-4, print; ISBN 952-462-291-2, online.

E:35 Katja Taipalus **Bubbles in the Finnish and US equities markets**. 2006. 123 p. ISBN 952-462-306-4, print; ISBN 952-462-307-2, online.

E:36 Laura Solanko **Essays on Russia's Economic Transition**. 2006. 133 p. ISBN 952-462-316-1, print; ISBN 952-462-317-X, online.

E:37 Mika Arola **Foreign capital and Finland Central government's first period of reliance on international financial markets, 1862–1938.** 2006. 249 p. ISBN 952-462-310-2, print; ISBN 952-462-311-0, online.

E:38 Heli Snellman **Automated Teller Machine network market structure and cash usage.** 2006. 105 p. ISBN 952-462-318-8, print; ISBN 952-462-319-6, online.

E:39 Harry Leinonen (ed.) **Simulation studies of liquidity needs, risks and efficiency in payment networks**. 2007. Proceedings from the Bank of Finland Payment and Settlement System Seminars 2005–2006. 320 p. ISBN 978-952-462-360-5, print; ISBN 978-952-462-361-2, online.

E:40 Maritta Paloviita **Dynamics of inflation expectations in the euro area**. 2008. 177 p. ISBN 978-952-462-472-5, print; ISBN 978-952-462-473-2, online.

E:41 Charlotta Grönqvist **Empirical studies on the private value of Finnish patents**. 2009. 162 p. ISBN 978-952-462-498-5, print; ISBN 978-952-462-499-2, online.

E:42 Harry Leinonen (ed.) **Simulation analyses and stress testing of payment networks**. 2009. Proceedings from the Bank of Finland Payment and Settlement System Seminars 2007–2008. 340 p. ISBN 978-952-462-512-8, print; ISBN 978-952-462-513-5, online.

ISBN 978-952-462-512-8
ISSN 1798-1077

Multiprint Ltd
Helsinki 2009