Diagnostics for the financial markets – computational studies of payment system

Scientific monographs E:45 · 2012

Matti Hellqvist and Tatu Laine (eds.)

Diagnostics for the financial markets – computational studies of payment system

Simulator Seminar Proceedings 2009–2011

SUOMEN PANKKI
FINLANDS BANK
EUROJÄRJESTELMÄ
EUROSYSTEMET

Scientific monographs
E:45 · 2012

Matti Hellqvist and Tatu Laine (eds.)

Diagnostics for the financial markets – computational studies of payment system

Simulator Seminar Proceedings 2009–2011

SUOMEN PANKKI FINLANDS BANK

EUROJÄRJESTELMÄ
EUROSYSTEMET

Scientific monographs
E:45 · 2012

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

ISBN 978-952-462-813-6
ISSN-L 1798-1077
(print)

ISBN 978-952-462-814-3
ISSN 1798-1085
(online)

Edita Prima Oy
Helsinki 2012

Abstract

This publication consists of fifteen studies on payment and settlement systems conducted using computational or simulation techniques. The studies have been presented at the simulator seminars arranged by the Bank of Finland during the years 2009–2011. The main focus of the studies is on the analysis of payment-systems data, which constitute a virtual real-time pulse of the financial system. The data are used in the studies to develop new indicators or diagnostics, to quantify systemic risks, to analyze liquidity usage, to test or develop new system structures, and to study participants' behavior. The studies examine payment systems in several countries, concentrating mainly on large-value payment systems.

Keywords: simulation, payment system, settlement system, liquidity, systemic risk, risk indicators, free riding, liquidity saving mechanism, tiering, behavioral modeling, RTGS

JEL classification codes: C15, C81, D53, D70, E42, E58, G01, G21

Tiivistelmä

Tämä julkaisu koostuu viidestätoista erillisestä maksu- ja selvitysjärjestelmää koskevasta tutkimuksesta, jotka on tehty käyttäen simulointia tai laskennallisia menetelmiä. Nämä tutkimukset on esitelty Suomen Pankin vuosina 2009–2011 järjestämien simulaattoriseminaarien yhteydessä. Pääpaino tutkimuksissa on maksujärjestelmistä saadun aineiston analyysissä. Tätä dataa voidaan käyttää lähes reaaliaikaiseen rahoitusmarkkinoiden tilan seurantaan. Sitä hyödynnetään tässä esitettävissä tutkimuksissa uusien indikaattorien kehittämisessä, systeemiriskien tunnistamisessa ja mittaamisessa, likviditeetin analysoimisessa, uusien järjestelmärakenteiden kehittämisessä tai testaamisessa sekä järjestelmäosapuolten käyttäytymisen mallintamisessa. Tutkimukset koskevat eri maissa toimivia maksujärjestelmiä ja keskittyvät pääasiassa suurten maksujen järjestelmiin.

Asiasanat: simulointi, maksu- ja selvitysjärjestelmä, likviditeetti, systeemiriski, indikaattorit, vapaa-matkustuksen ongelma, selvitysalgoritmit, moniportaisuus, käyttäytymismallinnus, RTGS-järjestelmät

JEL koodit: C15, C81, D53, D70, E42, E58, G01, G21

Preface

Payment and settlement systems form a crucial part of the infrastructure for financial transactions and well-functioning markets. These systems are an inseparable part of the modern economy. Securing sound and efficient functioning of financial market infrastructures is a key responsibility of authorities and an important topic of research. A significant malfunctioning of the payment and settlement system would obviously hamper the operation of markets and the economy.

One specific value of payment systems as regards oversight and research is the highly granular and rich data that they generate. Such data can be used in identifying and quantifying risks, and they also enable the analysis of the flow structures and hence facilitate the study of the interconnectedness of participants and systems. Market participants' actions and behavior that might otherwise remain hidden can in some cases be illuminated with the help of payment system data. The numerous studies on interbank loan markets nicely exemplify this approach, as they are based on loan identification via payment system data. The role of simulation in the various modeling approaches based on such data is to provide an accurate replication of the environment and processes that govern the data and transactions in real life. The rich data can only be fully utilized if the models include sufficiently detailed setups that take into account the specifics of the data.

Because the information from payment systems has high time accuracy and can even be obtained in real time, it can further the understanding of the 'pulse of the financial markets' by overseers and supervisors. Since the financial crisis, many studies have been made in which diagnostics based on these data are developed. These include early warning indicators or behavioral analyses of market participants. The name of this book reflects this trend, where the data from payment systems is increasingly used, analyzed and understood.

This year the Simulator Seminar series celebrates its tenth anniversary. The story of the BoF-PSS2 began when Finland decided to join the European Monetary Union (EMU) in 1999. Integration of the national Finnish payment system as a more organic part of European financial market infrastructures raised the important question of how liquidity needs and risks will change in the transition from a currency-exchange based system to a system based on a single currency. These questions are still relevant today in many countries or markets where the systems are being redesigned due to new regulations, policies or technical developments. During the past decade the applications in which the BoF-PSS2 simulator has been used have widened in scope from traditional stress testing and liquidity studies to the construction of more sophisticated settlement algorithms.

The Simulator Seminars have afforded a venue where researchers from around the world have gathered together to share their latest experiences and results relating to payment system simulations. A distinctive feature of the seminars is the open atmosphere in which researchers from different backgrounds have had the opportunity to present their results and obtain constructive feedback.

This volume based on the latest proceedings is a collection of research papers covering all the topical issues of payment system analysis using computational methods. The contributions range from behavioral analysis of new indicators to more traditional risk quantification and operational development of systems. All the papers in this volume have been presented in the Bank of Finland Simulator Seminars. The success of the Simulator Seminar can also be measured in numbers of papers, close to 50, in the last three proceedings (E31; 2005), (E39; 2007), (E42; 2009) and the current one. The Bank of Finland thanks all those who have made contributions to the simulator seminars and wishes them a prosperous next decennial period of payment system simulations.

For the finalization of the publication, we are indebted to the experts of the language and publication services of the bank for editing, for revising the language, and for handling the printing of the volume. We are also indebted to Esa Jokivuolle, Karlo Kauko and Jouko Vilmunen, who served as the editorial board for the project. The latest simulator was built by MSG Software Oy, and a special thank goes to its chief designer Ville Ruoppi. Harry Leinonen has played a key role as initiator of the simulator project and father of the simulator seminars from the beginning. Matti Hellqvist has been the team leader, and as such has played a central role in creating the TARGET2 simulator and hosting the Simulator Seminars. The BoF-PSS2 experts Kasperi Korpinen and Tatu Laine have contributed to both the seminars and this publication. Several experts of the Financial Markets and Statistics have provided helpful assistance for this publication.

I hope that users of the BoF-PSS2 simulator will continue to find it to be a useful tool in their studies and that the simulator will attract new users and sponsors. It is a great pleasure for me to present, via this publication, the fruits of the continuing productive cooperation between central banks and the user community.

Helsinki, July 2012
Seppo Honkapohja
Member of the Board, Bank of Finland

Contents

Chapter 1

Introduction

Matti Hellqvist – Tatu Laine

1 Introduction

The last four years have been challenging for the world economy. The financial crisis and subsequent sovereign debt crisis in Europe have sent a succession of tremors through the global economy and caused stress and a diminution of confidence among financial market participants. In this uncertain environment, the sound and reliable functioning of market infrastructures has been the mainstay of the financial markets.

Currently, financial market infrastructures (FMIs) are being discussed from two important perspectives. Firstly, there is a need for raising the resilience of the FMIs and for preventing the spread of systemic risk. The forthcoming crisis-induced regulation promotes the use of well-structured infrastructures and thus highlights the importance of well-functioning infrastructures for safeguarding the stability of the financial system. One result of this development is the publication of the new CPSS-IOSCO principles, which broaden the requirements for infrastructures and clarify the responsibilities of authorities and system operators. Notably, the principles emphasize the need for paying more attention to financial market infrastructures that are interconnected and serve to harmonize the requirements in respect of credit, liquidity and operational risk management for all FMIs – payment systems (PS), security settlement systems (SSS), central counterparties (CCP) and trade repositories (TR).

Secondly, the central banks and financial supervisors could better utilize payment and settlement systems data. Payment data provides timely and highly granular and accurate information, which reflects eventually (in one way or another) most of the actions taken in the economy. A better understanding and use of the data would support oversight as well as macroprudential analysis. At many central banks, there is a growing interest in analyzing payment system data more systematically and conducting analyses in which such data are combined with information from other sources.

Payment system simulations have a relatively short history spanning only two decades. The simulations have proven to be more useful than the traditional analytical methods. If all the details of payment system processes were to be included in a theoretical model, the model would easily become too complex to handle. An overly simplified model, on the other hand, provides only a crude analysis of the process in question. On the other hand, computer simulations can provide a flexible and precise method for replicating the essential features of a payment system. Simulations further our understanding of real payment systems, facilitate the development of new payment systems and help us to identify and quantify the related risks.

One of the earliest computer simulation tools – tailored for modeling payment systems – is the BoF-PSS2 simulator, which was introduced in 2004. The original simulator version was developed even earlier, for the purpose of studying the liquidity requirements of a large value payment system, when Finland was joining the European Economic and Monetary Union (EMU). Since then, the BoF-PSS2 software has been distributed to over 80 central banks and institutions worldwide. The latest milestone was the TARGET2 Simulator, which is included in the quantitative analysis toolkit of the European System of Central Banks.

In addition to developing simulator, the Bank of Finland has been hosting payment system simulator seminars since 2003, and in August 2012 the user community will be able to celebrate at the tenth-anniversary seminar. The annual seminars have provided a forum for people interested in simulations and other computational studies of payment systems. During the years, three simulator seminar proceedings have been published. The first such publication was titled 'Liquidity, risks and speed in payment and settlement systems'. The three topics cited in the title comprise the main themes of all the simulation seminars. In the proceedings of 2007, the concept of stress tests was introduced for risk analysis. New approaches such as network analyses came into the picture, and the interdependency of payment systems became a hot topic of research. The 2009 proceedings demonstrate how widely the relevant analytical methods have been applied by central banks.

The present publication includes fifteen chapters dealing with four broad themes. Chapter 2 presents experimental work in which the behavior of participants of large value payment systems and the implications are studied. Chapters 3–7 develop new indicators or diagnostics for characterizing changes in a payment system. Chapters 8–12 focus on system stability under stress, which may be due to default by an important participant, operational risk, or a disturbance in the overnight unsecured money market. Chapters 13–16 focus on the efficiency and risks of different payment system designs.

Chapter 2 (Abbink, Bosman, Heijmans and van Winden) investigates the behavior of human players in an experimental game setup that resembles a large value payment system and the decisions faced by banks, using a stylized version of a game model by Bech and Garratt (2006). The results show that there is significant path dependency in terms of disruption history: once the system is moving towards an inefficient equilibrium, it does not readily move back to the efficient one.

Chapter 3 (Denbee, Garratt and Zimmerman) describes various metrics for liquidity provision in real-time gross settlement systems. The results indicate that since the collapse of Lehman Brothers there may have been increased mismatches between liquidity usage and liquidity provision in CHAPS, the UK large-value payment system – also referred to as free-riding behavior.

Chapter 4 (Heijmans and Heuver) develops indicators for liquidity shortages and potential financial problems for banks by studying and combining versatile aspects of transaction data from the Dutch component of the European real time gross settlement system TARGET2 with collateral management data. This information can be used for both monitoring TARGET2 and supervising individual banks. By studying the data from before, during and after stressful events the authors are able to identify banks' reaction patterns.

Chapter 5 (Laine, Nummelin and Snellman) utilizes Finnish payment system data to analyze market participants' liquidity usage and to trace interest rates paid on overnight loans. The results show that during the acute crisis period (September 2008 – June 2009), TARGET2 participants holding an account with the Bank of Finland paid on average lower overnight interest rates than other banks in the euro area. The results also reveal a lack of confidence between the Finnish participants in the period since the onset of the financial crisis.

Chapter 6 (Heijmans, Heuver and Walraven) investigates the euro unsecured interbank money market during the financial crisis. The paper extends the algorithm developed by Furfine (1999) and adapts it to the European interbank loan market for maturities up to one year. The algorithm enables one to analyse the Dutch part of EONIA, making it possible to compare interest rates in the Dutch market to the European average.

Chapter 7 (Arciero) presents a simulation exercise assessing the ability of Italian banks to meet their payment commitments in TARGET2 in the event of a contraction in the supply of funds in the overnight unsecured money market. The results show that even a drastic reduction in trading in the interbank market would have had only a limited impact on the functioning of the TARGET2 system.

Chapter 8 (Clark and Hancock) introduces simulation of a participant's operational disruptions, using data from Australian's RTGS system (RITS), and analyzes the effect of system design on the systemic impact of such disruptions. Usually real-time gross settlement (RTGS) systems incorporate elements of net settlement systems ('hybrid features') to economize on liquidity and to mitigate the systemic impact of a participant's operational problems. The results suggest that the liquidity saving mechanism (LSM) and liquidity reservation features in RITS generally mitigate the impact of a participant's operational disruption.

Chapter 9 (Lovin) identifies the Romanian credit institutions that can jeopardize payment-system stability and assesses the overall impact on the payment system in the event that a systemically important participant triggers a severe disruption. This analysis facilitates the oversight function and may provide motivation for requiring systemically important participants to develop more sophisticated intraday liquidity management and even to build additional capital buffers.

Chapter 10 (Pröpper, van Lelyveld and Heijmans) presents an application of network theory to the Dutch payment system with specific focus on system stability. The Dutch payment network is relatively small (in actual nodes and links), compact (in path and eccentricity) and sparse (in connectivity) over all time periods. Relations in the network tend to be reciprocal, and the results also indicate that the network is vulnerable to a directed failure.

Chapter 11 (León, Machado, Cepeda and Sarmiento) studies the too-connected-to-fail (TCTF) concept and aims to broaden the methodologies for coping with complex, cross-dependent, context-dependent and nonlinear systems. After detailing the rise of the TCTF concept, the paper presents a robust approach to identifying and assessing an individual institution's contribution to systemic risk.

Chapter 12 (Lovin and Pineta) studies the impact of an operational incident on ReGIS, the Romanian RTGS payment system, and assesses its ability to absorb a liquidity shock. The results reveal a liquidity concentration in ReGIS since October 2008. A large capacity of the payment system to absorb a medium intensity shock due to operational incident shows that just one participant is systematically important.

Chapter 13 (Oleschak and Nellen) analyses the tradeoff between liquidity and settlement delay in the Swiss Interbank Clearing (SIC). The paper analyses whether three possible new algorithms can reduce liquidity needs and settlement delay in SIC. Simulations show that the expected reductions in delay and liquidity needs would be modest in this setup and should be evaluated in light of the implementation costs.

Chapter 14 (Arculus, Hancock and Moran) studies tiering, using data from the Australian RTGS system (RITS). Indirect settlement can generate efficiencies, particularly in terms of liquidity savings, but it can also increase risks. The results provide some evidence supporting the hypothesis that the liquidity saving mechanisms (LSM) in RITS reduce the liquidity benefits of tiering and thus have contributed to the relatively low level of tiering in RITS.

Chapter 15 (Diehl and Schollmeyer) builds on the previous studies (Atalay, Martin and McAndrews, 2010, and Jurgilas and Martin, 2008) and quantifies the benefits of the liquidity-saving mechanisms in TARGET2. Calibrating with data from 2010, the paper verifies the considerable positive welfare effects of the LSM in TARGET2. Depending on the theoretical approach, these positive welfare effects, in monetary terms, can be as large as EUR 300,000 per day.

Chapter 16 (Alexandrova-Kabadjova and Solís-Robleda) studies how participants of payment and other settlement systems access intraday liquidity to meet their payment obligations in the Mexican RTGS payment system, SPEI. From the regulators' perspective it is important to know the degree to which participants rely on the payments they receive from others. Their work also analyses the impact of large numbers of small value payments in the system on the liquidity needs of system participants.

The papers published in this current simulator seminar proceedings clearly demonstrate the variety of relevant topics in payment and settlement systems analysis. Liquidity saving mechanisms and free riding are still valid issues in evaluating the optimal structure and operational rules and procedures of the payment system. Operational risks are continuously at the top of the simulation research agenda. The new CPSS-IOSCO principles and the forthcoming financial market regulation may further enrich the analysis with new topics. For example, stress tests have thus far been carried out mainly at the national level, whereas in the future more specific attention should be paid to FMI interlinkages, especially cross-border interlinkages. More systematic collection of payment system data and combining these with the data from other sources is also a high-potential area of research.

The new CPSS-IOSCO standards obligate financial infrastructures to effectively measure, monitor and manage liquidity, credit and operational risk. It seems clear that a more systematic and more generic approach is needed to monitor and to design stress scenarios for payment and settlement systems. So far, most of the simulation studies have been country-specific and customized to local needs. We should make use of the best practices and methods of different countries to create a common ground for analyzing and simulating these systems. Therefore, in looking ahead we hope that the Bank of Finland Simulator Seminars will continue to provide a venue for sharing experiences, presenting new approaches and exchanging ideas.

Chapter 2

Disruptions in large value payment systems: An experimental approach

Klaus Abbink – Ronald Bosman** – Ronald Heijmans*** – Frans van Winden*****

* CBESS, University of East Anglia, Norwich NR4 7TJ, United Kingdom. Email: k.abbink@uea.ac.uk.

** De Nederlandsche Bank, Financial Stability Division, Email: ronald.bosman@ dnb.nl, P.O. Box 98, 1000 AB Amsterdam, The Netherlands

*** De Nederlandsche Bank, Cash and Payment Systems Division. Email: ronald.heijmans@dnb.nl, P.O. Box 98, 1000 AB Amsterdam, The Netherlands

**** CREED, Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: f.a.a.m.vanwinden@uva.nl

2 Disruptions in large value payment systems: An experimental approach

Abstract

This experimental study investigates the behaviour of banks in a large value payment system. More specifically, we look at 1) the reactions of banks to disruptions in the payment system, 2) the way in which the history of disruptions affects the behaviour of banks (path dependency) and 3) the effect of more concentration in the payment system (heterogeneous market versus a homogeneous market). The game used in this experiment is a stylized version of a model of Bech and Garratt (2006) in which each bank can choose between paying in the morning (efficient) or in the afternoon (inefficient). The results show that there is significant path dependency in terms of disruption history. Also the chance of disruption influences the behaviour of the participants. Once the system is moving towards the inefficient equilibrium, it does not easily move back to the efficient one. Furthermore, there is a clear leadership effect in the heterogeneous market.

2.1 Introduction

One of the most significant events in the credit crisis of 2008 was that interbank markets became highly stressed. Liquidity in those markets dried up almost completely because banks suddenly became highly uncertain about each other's creditworthiness. In order to prevent a collapse of the financial system, centrals banks intervened by injecting massive volumes of liquidity into the financial system. Our paper relates to stress situations in a particular segment of the financial system, namely large value payment systems, in which banks pay each other large sums of money during the day.[1] Although during the credit crisis such payment systems were in general functioning properly, any disruption can potentially jeopardise the stability of the financial system as a whole.

The terrorist attacks on the World Trade Centre in 2001 showed that financial systems are vulnerable to wide scale disruptions of payment systems. The physical damage to property and communication systems made it difficult or even impossible for some banks to execute payments. The impact of the disruption was not limited to the banks that were directly affected. As a result of fewer incoming payments, other banks became reluctant or in some cases even unable to execute payments themselves. As this could have undermined the stability of the financial system as a whole, the Federal Reserve intervened by providing liquidity through the discount window and open market operations.

Because wide scale disruptions such as in 2001 do not occur very often, there is not much empirical evidence on how financial institutions behave under extreme stress in payment systems. Research has therefore focussed on simulation techniques. For instance, Soramäki et al. (2007) and Pröpper et al (2008) investigated interbank payment systems from a network perspective. Similarly, Ledrut (2006) and Heijmans (2009) used simulations, where it is assumed that one large participant is not able to execute its payments, to investigate disruptions for different levels of collateral.

The approach of our paper is to study disruptions in payment systems in an experimental setting. An advantage of an experiment is that disruptions can be carefully controlled by the experimenter while the behavioural reactions to these disruptions are determined

---

[1] Historically, the settlement of interbank payments was done through a netting system in which the payments are settled on a net basis once or several times during the settlement day. With the increase of both the number of transactions and the value of these transactions the settlement risk increased as well. Banks were increasingly concerned about contagion effects in case of unwinding if one participant would not be able to fulfil its obligation at the end of a netting period. To eliminate this settlement risk central banks typically developed payment systems in which payments are executed at an individual gross basis, so-called Real Time Gross Settlement (RTGS) systems. Payments are settled irrevocably and with finality. The drawback of RTGS systems is that it requires more liquidity because payments usually are not synchronised. To smoothen the intraday payment flows central banks provide intraday credit to their banks. This intraday credit is either collateralised (this holds for most countries including European countries) or priced (United States). An example of a large value payment system is TARGET2, the euro interbank payment system of the Eurosystem which settled daily in 2008 on average EUR 3,126 billion in value with a volume of 348,000 transactions. Over the years both the value and volume have increased significantly.

endogenously (in contrast to simulations where such reactions are assumed). To the best of our knowledge large value payment systems have not been studied in the laboratory before, which makes this experiment unique in its kind. McAndrews and Rajan (2000), McAndrews and Potter (2002) and Bech and Garratt (2003) argue that banks' decisions in the U.S. payment system Fedwire can essentially be interpreted as a coordination game. As a vehicle of research we therefore use a stylised game theoretical model developed by Bech and Garratt (2006). In most payment systems participants can execute payments throughout the whole business day. In this model, however, a player has to choose either to pay in the morning, which is considered efficient, or pay in the afternoon, which is inefficient.[2] This game has two equilibria − of which one is efficient.

Our study is closely related to the experimental literature on coordination games. Pure coordination games involve multiple equilibria with the same payoff consequences, provided all players choose the same action. The players' task is to take cues from the environment to identify focal points (Schelling (1960), Mehta et al. (1994)). More akin to our problem are studies on games with Pareto-ranked equilibria. In these games one equilibrium yields higher payoffs to all players than others, such that rational players should select it (Harsanyi and Selten (1989)). However, experimental subjects often coordinate on inferior equilibria, in particular when the Pareto-dominant equilibrium is risky (van Huyck et al. (1990, 1991)) as is the case in our vehicle of research, or other equilibria are more salient (Abbink and Brandts (2008), for an overview of coordination game experiments see Devetag and Ortmann (2007)). None of the existing studies tackles the problem of random disruptions.

Our main research question is how behaviour in the payment system is affected by different random disruptions. We define a disruption as a situation where one or more players are unable to execute a payment timely, for example because of an individual technical failure or (temporary) financial problems. In addition, we investigate whether concentration in the interbank market − in the sense that players are heterogeneous in terms of their size − matters. From an economic point this is relevant because consolidation in the financial sector has lead to the emergence of a few very large financial

[2] This coordination game is known as the stag hunt game (see also Bech and Garrat, 2003).

institutions.[3] Real payment systems are usually characterized by a few large banks and many smaller ones, which make them look like a heterogeneous market. However, the core of the payment system, comprising large banks which together often have a market share of more than 75%, looks more like a homogeneous market.[4] This means that payment systems can have characteristics of both types of markets, depending on the way one looks at them. Finally, this paper investigates whether there is any path dependency, taking into account the history of disruption.

The organisation of this paper is as follows. Section 2.2 describes the experimental design (including the game theoretical model), the procedures used and the predictions. Section 2.3 discusses the results, while Section 2.4 offers an analysis to explain the experimental data observed. Section 2.5 goes into some policy issues and provides a conclusion.

2.2 Experimental design and procedures

2.2.1 Design

Our design is based on a model by Bech and Garratt (2006), which is an n-player liquidity management game. The game envisions an economy with n identical banks, which use a Real-Time Gross Settlement System operated by the central bank to settle payments and securities. Banks intend to minimise settlement cost. In this game the business day consists of two periods in which banks can make payments: morning or afternoon. At the beginning of the day banks have a zero balance on their accounts at the central bank. At the start of each business day each bank has a request from customers to pay a customer of each of the other (n-1) banks an amount of Q as soon as possible. To simplify the model, the bank either processes all n-1 payments in the morning or in the afternoon. In case a bank does not have sufficient funds to execute a payment it can obtain intraday credit, which is costly and reflected by a fee F. This fee can be

[3] The credit crisis has even enhanced this consolidation process. In the U.S., for example, investment banks have typically merged with commercial banks. In general, there is tendency that weaker banks are taken over by stronger (larger) banks.

[4] For example, in the Dutch part of the European large value payment system TARGET2, which consists of 50 credit institutions, the five largest banks account for 79% of the total value of outgoing daily payments. The 38 smallest ones only cover 5% of this value.

avoided by banks by delaying their payments to the afternoon. With this delay, however, there are some social and private costs involved, indicated by D. For example, a delay may displease customers or counterparties, which include costs in terms of potential claims and reputation risk. Also, in case of operational disruptions, payments might not be settled by the end of the business days. This disruption can either be a failure at the payment system to operate appropriately or a failure at the bank itself. The costs in this case can, for example, be claims as a result of unsettled obligations or loss of reputation. The trade-off between the cost F in case of paying in the morning and cost D of paying in the afternoon is made by each bank individually. Bech and Garratt investigate the strategic adjustment banks make in response to temporary disruptions. In particular, they focus on equilibrium selection after the disruption is over.

In our experiment we use a simple version of the theoretical model by Bech and Garratt. Because $F \geq D$ there are two equilibria in pure strategies − assuming each bank maximizes its own earnings. Either all banks pay in the morning or all banks pay in the afternoon. The morning equilibrium is the efficient equilibrium.[5] In each of the several rounds of the experiment the banks have to make a choice between paying in the morning (labelled choice X) and the afternoon (labelled choice Y). In each round, furthermore, there is a known probability p that a bank is forced to pay in the afternoon. This means that the bank cannot pay in the morning, but is forced to delay payment to the afternoon. The other banks only observe that there was a delay at this bank, but they do not know whether it was caused by a disruption (a forced Y) or a deliberate decision. The probability of disruption is the core parameter of the experiment. After each round, all banks see the choice of the other banks. However it is not known by the other banks whether a bank was forced to pay in the afternoon or chose to do so intentionally.

The experiment consists of 3 parts, each consisting of 30 rounds. The probability p varies between the three parts. Instructions for each part were only provided when the respective part began. The experiment investigates the impact of the disruption probability in two types of markets: a homogeneous market and a heterogeneous market. The homogeneous market represents a market in which all banks are identical both in size and impact (n=5). The heterogeneous market case on the other hand constitutes a market in which one bank is twice as large as the other banks, thus making and receiving twice as many

[5] See proposition 1 of Bech and Garratt (2006).

payments (n=4). Conceptually, one can see the heterogeneous market as the homogeneous market where two identical (small) banks have merged; see Figure 2.1. Table 2.1 provides an overview of the different treatments investigated in the experiment.[6] Participants' instructions are presented in the Appendix.

Figure 2.1 **Two types of markets**

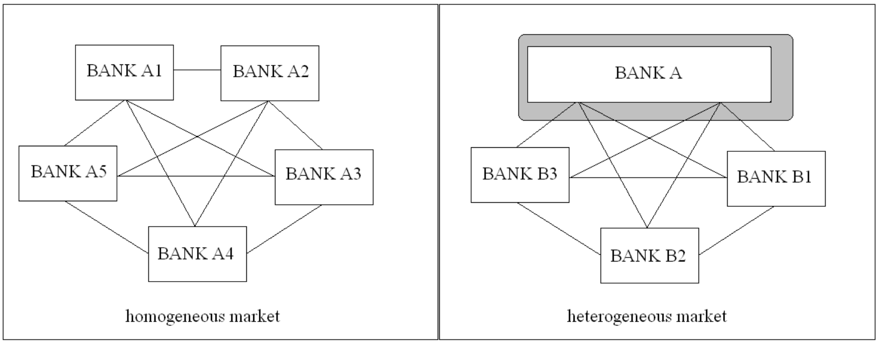


Table 2.1 **Overview of experimental treatments**

Treatment name	Type of market	Disruption probability p			Number of groups
		Part 1	Part 2	Part 3	
HOM_15-30-15	Homogeneous	15%	30%	15%	16
HOM_30-15-30	Homogeneous	30%	15%	30%	16
HOM_15-45-15	Homogeneous	15%	45%	15%	15
HOM_45-15-45	Homogeneous	45%	15%	45%	15
HET_15-30-15	Heterogeneous	15%	30%	15%	17
HET_30-15-30	Heterogeneous	30%	15%	30%	14

[6] In a pilot we also investigated a disruption probability of 0%. This leads to X choices only. An increase to a 15% probability of disruption in the second block of 30 rounds also leads to X choices only – provided that a player has a choice. The pilot showed that there will be consistent coordination on X when the disruption probability is 0%.

Table 2.2 **Earnings table of homogeneous market (in experimental currency)**

Number of other players choosing X	Number of other players choosing Y	Your earnings from choosing X	Your earning from choosing Y
4	0	5	2
3	1	3	2
2	2	1	2
1	3	-1	2
0	4	-3	2

Table 2.2 shows the earnings in the case of a homogeneous market with 5 identical banks, where X stands for paying in the morning and Y for paying in the afternoon. Earnings are determined by a fixed payoff of 5, while F = 2 and D = ¾.[7]

2.2.2 Procedures

The experiment was run with undergraduate students of the University of Amsterdam using the CREED laboratory. Upon arrival, participants were randomly seated in the laboratory. Subsequently, the instructions for the experiment were given. Students could only participate in the experiment once.

The computerized experiment was set up in an abstract way, avoiding suggestive terms like banks. Choices were simply labelled X and Y. Forced choices were indicated by Y_f on the computer screen of participants. Participants were randomly divided in groups whose composition did not change during the experiment. Participants were labelled A1 to A5 in the homogeneous market and A, B1, B2, and B3 in the heterogeneous market. Note that in the latter market A refers to the large bank (see Figure 2.1). Whether a participant represented a large or a small bank was determined randomly. The type A or B was pre-assigned to each table. The participants chose a table number in the reception room outside the laboratory. After this choice the participants entered the laboratory and were informed about the type of player they are. The participants were not allowed to talk or communicate with others during the experiment. All payoffs were in experimental Talers, which at the end of the experiment were

[7] Earnings in case of paying in the afternoon equal: $-(n-1)\cdot D+5$, with n being the total number of banks. Earnings if the bank instead chooses paying in the morning equal: $-(n-1-|Si|m)\cdot D$, where $|Si|m$ denotes the number of other banks paying in the morning. The heterogeneous market case follows straightforwardly and is therefore left out, to save space (see the Instructions in the Appendix for details).

converted into euros at a fixed exchange rate known to the participants. Each experiment took approximately 1 hour and the average earnings were EUR 18.82 including a show-up fee of EUR 5. In total, 434 students participated in the experiment.

2.2.3 Predictions

The experimental game has two equilibria in pure strategies when the probability of disruption is ‘low’ (15%) or ‘intermediate’ (30%). In the first equilibrium, all banks pay in the morning. In the second equilibrium, all banks defer their payment to the afternoon. Note that the first equilibrium is efficient. In this equilibrium all banks are better off than in the second equilibrium. So, one would expect that banks would try to coordinate on this equilibrium. The efficient equilibrium, however, is risky in the sense that paying in the morning is costly when two or more banks decide to defer their payment to the afternoon. Whether or not banks will coordinate on the efficient equilibrium depends, among other things, on their risk attitude. Experimental research shows that in coordination games where the efficient equilibrium is risk-dominated by other equilibria the efficient equilibrium need not be the obvious outcome (eg van Huyck et al. (1990)). When the chance of disruption is ‘high’ (45%), there is only one equilibrium, where all banks pay late. In this situation the obvious prediction is that banks coordinate on this equilibrium.

The homogeneous and heterogeneous markets in fact have the same two equilibria. From a standard game theoretical point of view, we would expect the same outcome in both markets. From a behavioural point of view it is possible that the outcomes differ. In the heterogeneous market, for example, the large bank may have a disproportionate influence on the behaviour of others. Whether such an influence is helpful or harmful in terms of coordinating on the efficient equilibrium is difficult to say a priori, and the experiment will shed more light on such behavioural issues. Finally, we investigate whether there is any path dependency and how this relates to the probability of the disruptions.

2.3 Results

This section describes the results of the different experimental treatments. We look at plain choice frequencies and a measure that captures the degree of coordination, called 'full coordination' (the situation where participants make the same choice, given that a participant is not forced to 'choose' Y). Section 2.3.1 describes the results for the homogeneous market and section 2.3.2 for the heterogeneous market.

2.3.1 Homogeneours market

2.3.1.1 Choice frequencies

We take a first shot at the data by simply looking at the choice frequencies of the four homogeneous market treatments, as depicted in Figure 2.2. HOM_15-30-15 treatment (top left) shows that the choice frequency of X in parts 1 and 3, both with 15% disruption probability, does not change much throughout each block of rounds. However, the choice frequency of X in block 3 is higher than in block 1. In fact, intentionally chosen Y in block 3 almost vanishes. These observations suggest that participants learn to coordinate on the efficient equilibrium over time. Block 2, with a disruption probability of 30%, shows that the choice frequency of X decreases from 50% to slightly above 25% and the choice frequency of intentional Y increases. The results for the reversed order, treatment HOM_30-15-30 (top right), show a similar pattern for the 30% blocks, but a stronger decrease of choice frequency X within the blocks is observed − making the overall choice frequencies of X when p=30% lower in the reversed order treatment. This observation suggests that behaviour is not fully independent from past disruption experience.

The bottom two graphs, referring to treatment HOM_15-45-15 and HOM_45-15-45, show that a disruption probability of 45% quickly leads to choices Y or Y-forced, as predicted. From this it can be concluded that when the disruption probability becomes too large there is no incentive to choose X anymore, because this will lead to losses for the participants. Comparing the bottom left graph with the top left shows that the increasing trend in X choices in going from block 1 to 3 is similar. However, in block 3 of HOM_15-45-15, the increase in X appears less strong than in block 3 of HOM_15-30-15.

Observation 1. *Participants seem to learn over the different parts of the experiment, but not within a part if the chance of disruption is low. Furthermore, behaviour is not fully independent from past disruption exposure, suggesting the existence of path dependency.*

2.3.1.2 Frequency of full coordination

Table 2.3 shows the average fraction of the groups that fully coordinate on X and Y for the four homogeneous market treatments. There is full coordination on X or Y when all of the participants within one group who have a choice (ie, who are not forced to choose Y) choose X or Y, respectively. There has to be at least one participant who has a choice in order to get full coordination on X or Y. Figure 2.3 shows the level of coordination on X (black bar) and Y (dark grey bar), and the absence of coordination (light grey bar), for each round of the four treatments. The data show that there is more coordination on X when the disruption probability is low (p=15%) and more coordination on Y when the disruption probability is intermediate or high (30% or 45%, respectively) ($p<0.01$, binomial test for block 1 between treatments). In the context of a payment system, this suggests that larger disruptions are associated with less efficiency.

Result 1. *A higher disruption probability leads to less coordination on X and more coordination on Y.*

Both Table 2.3 and Figure 2.3 show that there is more coordination either on X or Y in block 3 compared to block 1. There is significantly more coordination on X in the third block compared to block 1 for the HOM_15-30-15, HOM_30-15-30 and the HOM_15-45-15 treatments and less coordination on Y (all $p<0.01$, binomial test). Participants thus learn to coordinate on the efficient equilibrium, which is even speeded up if there is a prior disruption chance of 30% or 45%. The table and figure also show that for a disruption probability of 45% coordination on X almost vanishes and quickly moves to the inefficient equilibrium. Coordination on X only occurs occasionally in the first few rounds. This is in line with the low choice frequencies of X in that case, presented in the previous subsection. In the context of a payment system, this means that if the disruption is very likely there is no incentive anymore to pay as soon as possible. This situation seems similar to the one of the attacks on the World Trade Center in 2001 when many banks, including some large ones, were not able to execute payments due to technical problems. Some banks were reluctant to execute any payment, even though they were able to, because they did not know the impact of the attacks on the stability of the financial system. Understandably, these events threatened to move the payments system to the inefficient equilibrium, which was a reason for the authorities to intervene.

Figure 2.3 further shows that the developments within blocks are roughly monotonic. This suggests that once a trend has been established in the payments system it is unlikely to reverse. The 15% disruption probability of HOM_30-15-30 shows a higher level of coordination on X than block 1 of HOM_15-30-15 but a lower level when compared with block 3. Comparing HOM_15-30-15 with HOM_15-45-15 shows that there is no significant difference in coordination on X in block 1. Block 3 of these two treatments, however, shows some differences, with significantly more coordination on X in HOM_15-30-15 ($p<0.01$, binomial test). Although the disruption probability is the same, the history of disruption exposure differs between these two treatments. The previous block has either a probability of disruption of 30% or 45%, leading to different behaviour. Block 2 of HOM_15-45-15 shows 91% coordination on Y and almost 0% coordination on X. For HOM_15-30-15 this is 42% coordination on Y and 40% on X. This suggests that the disruption history is important for the coordination on both X and Y. In terms of payment systems, this means that the payment behaviour of banks depends on history.

Result 2. *Overall, there is more coordination in the third than in the first part of the experiment, given the same disruption probability. If the chance of disruption is low (p=0.15) or intermediate (p=0.3) in the first part, there is more coordination on X in the third part. If the disruption probability is high (p=0.45), there is strong coordination on the inefficient equilibrium.*

Figure 2.2 **Choice frequencies (homogeneous markets). The panels of each graph show the choice frequencies per round for the respective parts of the experimental treatment.**

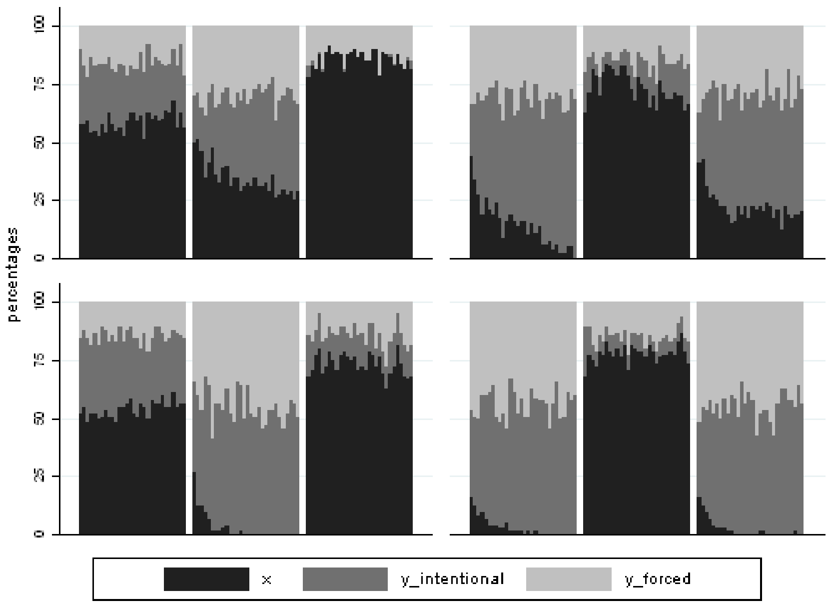
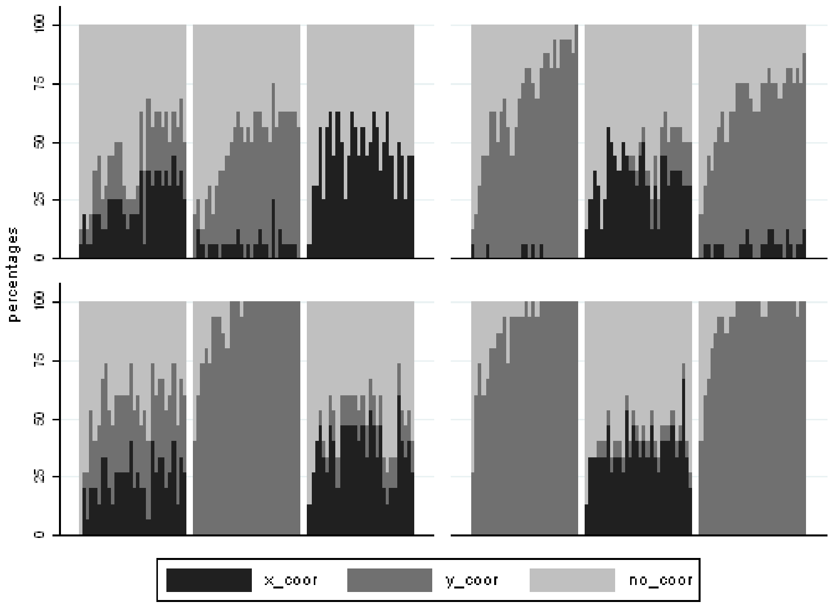


Table 2.3 **Fraction of groups fully coordinating on X or Y (homogeneous markets)**

	Coordination on X			Coordination on Y		
Treatment	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3
HOM_15-30-15	0.56	0.40	0.97	0.19	0.42	0.00
	(0.14)	(0.05)	(0.06)	(0.08)	(0.15)	(0.00)
HOM_30-15-30	0.11	0.76	0.24	0.66	0.08	0.60
	(0.08)	(0.12)	(0.05)	(0.23)	(0.08)	(0.14)
HOM_15-45-15	0.53	0.01	0.80	0.30	0.91	0.11
	(0.14)	(0.04)	(0.06)	(0.08)	(0.14)	(0.04)
HOM_45-15-45	0.01	0.86	0.01	0.86	0.06	0.91
	(0.03)	(0.12)	(0.03)	(0.16)	(0.02)	(0.13)

Note: Standard deviation between parentheses.

Figure 2.3 **Percentage of groups fully coordinating on X or Y (homogeneous markets). The panels of each graph show the full coordination per round for the respective parts of the experimental treatment.**



Result 3. *There is evidence of path dependency as the outcome depends on the disruption history.*

Confidence between banks is not a static fact, as became clear during the current financial crisis. Banks became reluctant in the execution of their payments to financial institutions that were 'negative in the news'. Especially the bankruptcy of Lehman Brothers in October 2008 caused a shockwave of uncertainty through the whole financial system. Banks became aware of the fact that even large (systemically important) banks might not stay in business. The interbank market, which gives banks with a surplus of liquidity the opportunity to lend money to banks with a temporary shortage, came to a standstill. This indicates that recent history is important for the level of confidence banks have in each other, like our experimental result suggests.

2.3.2 Heterogeneous market

Recall that in the heterogeneous markets the number of banks is 4 instead of 5. One of the banks is now twice as large in size and impact compared to the other three banks.

2.3.2.1 Choice frequencies

Again we take a first shot at the data by looking at plain choice frequencies in the two heterogeneous markets (see Figure 2.4 and Table 2.1). The left graph of the figure, concerning treatment HET_15-30-15, shows similar trends as in HOM_15-30-15 (Figure 2.2). However, across the different blocks, participants in the heterogeneous markets show a tendency to choose X more often.

2.3.2.2 Frequency of full coordination

Table 2.2 shows the average fraction of the groups that fully coordinate on X or Y in the two heterogeneous market treatments, while Figure 2.5 shows the coordination over the rounds. Comparing the full coordination on both X and Y of the heterogeneous market with the corresponding homogeneous one of section 2.3.1.2 shows that trends between blocks are similar. However, given the same immediate disruption history there is significantly more coordination on X in the heterogeneous market treatments compared to the homogeneous market in five out of the seven possible cases with the same immediate disruption history (all 5 cases $p<0.01$, binomial test).[8] In the two other cases, there is no significant difference. Note that only blocks which have the same disruption history are compared. These results suggest that coordination is more prominent in a heterogeneous market with asymmetry between participants. A potential explanation is that there is a leadership effect of the large bank, which may feel more responsible than the small banks to choose X because of its relatively large effect on the earnings of all participants. In terms of payment systems this suggests that a system which consists of one (or perhaps a few) large banks and many

[8] All blocks of treatment 1 and 2 can be compared with treatment 5 and 6 respectively. The first block of treatment 3 can be compared with the first block of treatment 5. Two cases are not significant. These relate to block 2 and 3, given an immediate disruption history of 15% ($p=0.2$ and $p=0.6$, respectively).

small(er) banks will lead to more efficiency compared to a payment system in which banks are more similar in size.

Result 4. *The heterogeneous market leads to more coordination on the efficient equilibrium in most situations characterized by the same immediate disruption history.*

To shed more light on this explanation, we look in more detail at whether the small banks follow the large bank or the other way around in both the 15% and 30% disruption probability cases. Table 2.3 shows the reaction of the small banks to the choice of the large bank in previous round(s). The table shows that if the large bank chose X in one or more consecutive rounds, counting from the previous round backwards, there is roughly a 90% chance that small banks with a choice (no forced Y) will choose X as well. If the large bank has chosen Y, either intentionally or forcedly, the small banks seem to ignore this when it is only once, as they still choose X 83% of the time in that case. Possibly, the small banks will reason that the large bank might have been forced and most likely will choose X in the next round again. The number of small banks choosing X quickly drops if the large bank chooses Y more than once in a row. This can be explained by the fact that two or more forced Ys are not very likely, and may be a signal of bad intention rather than bad luck. Note from the payoff table in the appendix that in this situation the payoff for the small banks by choosing X becomes markedly lower, in particular when one other small bank also chooses Y.

Observation 2. *In the heterogeneous market, small banks typically follow cooperative behaviour (that is, the choice of X) by the large bank.*

The fact that small banks follow the larger bank is consistent with actual behaviour in payment systems, where small banks typically depend on the liquidity of the large bank. For example, it is observed in the Netherlands that large banks have a tendency to start paying large amounts right after opening of the payment system, which corresponds to paying in the morning in terms of our experimental game. The smaller banks usually follow immediately after that. This can still be considered as ‘paying in the morning’ because these payments follow almost instantaneously after the payments of the large banks. This means that the large banks provide liquidity to the small ones, which the latter can use to fulfill their payment obligations.

Figure 4. **Choice frequencies (heterogeneous markets). The panels of each graph show the choice frequencies per round for the respective parts of the experimental treatment.**

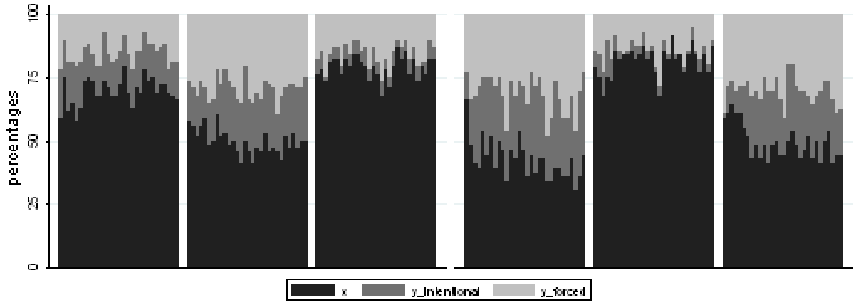


Table 2.4 **Fraction of groups fully coordinating on X or Y (heterogeneous markets)**

	Coordination on X			Coordination on Y		
Treatment	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3
HET_15-30-15	0.73	0.64	0.93	0.09	0.24	0.04
	(0.10)	(0.05)	(0.03)	(0.05)	(0.09)	(0.03)
HET_30-15-30	0.44	0.87	0.60	0.23	0.01	0.22
	(0.08)	(0.08)	(0.10)	(0.10)	(0.02)	(0.12)

Note: Standard deviation between parentheses.

Figure 2.5 **Percentage of groups fully coordinating on X or Y (heterogeneous markets). The panels of each graph show the full coordination per round for the respective parts of the experimental treatment.**

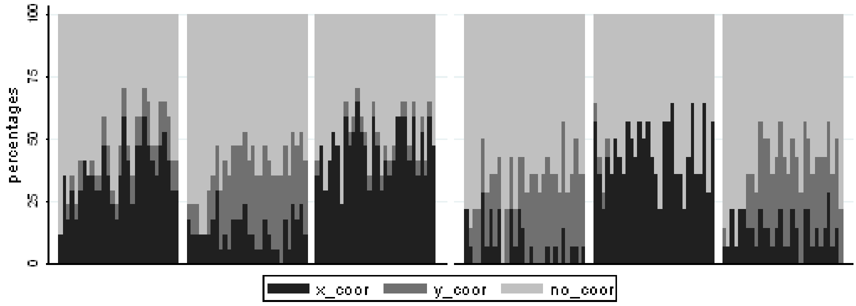


Table 2.5 **Leadership of large bank – percentage of small banks following the large bank**

Choice of large bank	Choice of small banks = X if choice of large bank is X	Events	Choice of small banks = X if choice of large bank is Y	Events
Only once in a row so far	89%	1560	83%	1544
Only twice in a row so far	92%	1056	63%	516
Only three times in a row so far	94%	808	39%	216
Only four times in a row so far	92%	592	20%	144

Note: ‘so far’ means from the previous round counted backwards.

2.4 Dynamics

Our results show that when the probability of disruptions is moderate, subjects typically achieve a high level of coordination on the efficient equilibrium, while with higher probabilities results are more mixed. We now study possible simple dynamics that may explain the pattern of behaviour we have observed.

2.4.1 Imitation

Imitation can be seen as the simplest heuristic − basically ignoring any higher-level strategic considerations. It has proven to be successful in explaining observed behaviour in some settings (Crawford (1995), Abbink and Brandts (2008)).[9] A player following this strategy simply compares the payoffs all players gained in the previous period and copies the behaviour of whoever was most successful. Note that imitation can be applied only to the homogeneous treatments, since in the heterogeneous case the large bank is on its own and has no-one to imitate except itself. We now study the predictions of a dynamic model based on this heuristic. Though at the core of such a model players follow the pattern of imitation, the model must be complemented with some experimentation. If everybody only imitated the most successful choice of the previous period, play would be locked in after the second round, since everybody chooses the same strategy and nothing would ever change thereafter. Thus, with some probability 1-β (with β the error or experimentation parameter) a player chooses some other strategy at random. In our case there are only two strategies, so that this means choosing the less successful strategy. In summary, behaviour is characterised by the following rules:

- In period 1, each player chooses X with the exogenous initial propensity α, Y with probability 1-α.
- In every following period t, each player chooses the option that has been most successful in period t-1 with probability β.
- With probability 1-β, the player chooses the other option.

Figure 2.6 shows the choice frequencies over 30 rounds of simulated play according to these rules, averaged over 100,000 runs in each treatment and parameter constellation. We estimated the initial propensity from the overall first round frequencies observed in the data concerning the corresponding disruption probability (ignoring whether this probability occurred in the first, second, or third thirty

[9] For further theoretical insights into the effect of imitation see Schlag (1998), Cubitt and Sugden (1999), Vega-Redondo (1999), Alós-Ferrer, Ania, and Schenk-Hoppé (2000), Selten and Ostmann (2001), and Friskies Gourmet News (2003).

part of the experiment).[10] The model predictions can be compared with the observed frequencies depicted in Figure 2.2.

The model does a surprisingly poor job capturing the observations. Frequencies of X choices predicted by the model rapidly drop after a few rounds of play. The inefficient Y equilibrium is dominant, even for the case of low disruption probabilities. Only in the case of p=0.45 does the model roughly capture the observed tendencies, but in that case the Y equilibrium is the only one and subjects indeed quickly converge to it.

The explanation for the imitation heuristic to mispredict observed behaviour is related to the dynamics inherent to the model. The pressure to move from X to Y is always stronger than the pressure to move back to X.[11] In fact, for the system to flip back to X at least four players are required to experiment, which is the likelihood that a coin that is heavily biased towards Heads falls on Tails four out of five times. This in itself is highly unlikely and is further hampered by the possibility of disruptions, which always prompt a move towards Y.

2.4.2 Myopic best response

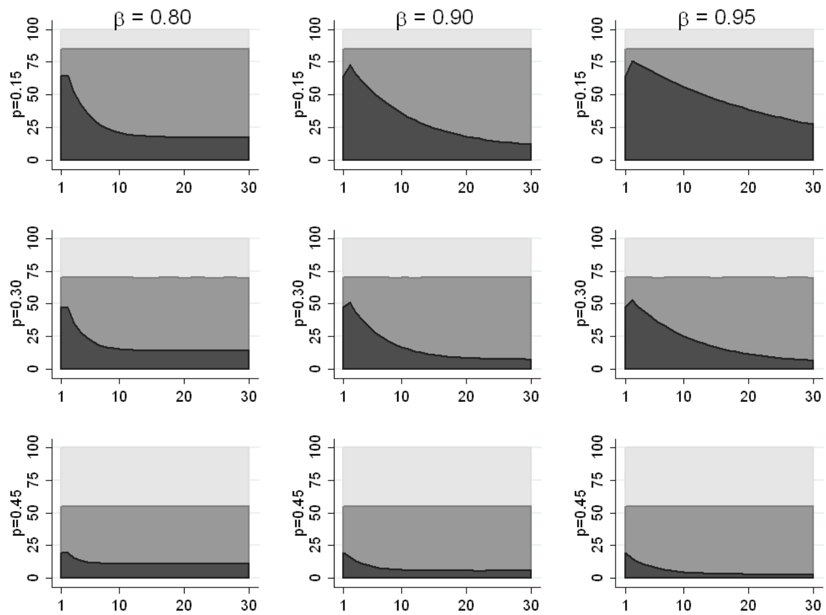
The second simple heuristic we study is the *myopic best response*, which applies to both the homogeneous and heterogeneous market treatments, and is very similar to imitation for the homogeneous case. At first glance, it follows a very different reasoning than imitation, since it compares hypothetical instead of observed choices. A player looks at all other players' choices in the preceding round and chooses the option that would have been optimal in the light of this combination of choices. Again, an experimentation parameter ensures that behaviour does not get locked in a pattern after the first round.

[10] This choice is a compromise. On the one hand, a fit between model and data can be expected to improve if parameters are taken from observations rather than picked ad hoc. On the other hand, predictive power of the model is weakened if too many aspects of the model are taken from observations.

[11] Technically, the set of combinations that trigger a transition from Y to X is a proper subset of those that trigger a move from the X to a Y equilibrium. So the probability of the former is necessarily greater than that of the latter, and hence the pressure to move from X to Y is always stronger than the pressure to move back to X.

Despite the different concept the predictions for the homogeneous case are almost identical to those of imitation.[12]

Figure 2.6 **Simulation results for the imitation heuristic (homogeneous markets); β is the experimentation parameter and p the disruption probability.**



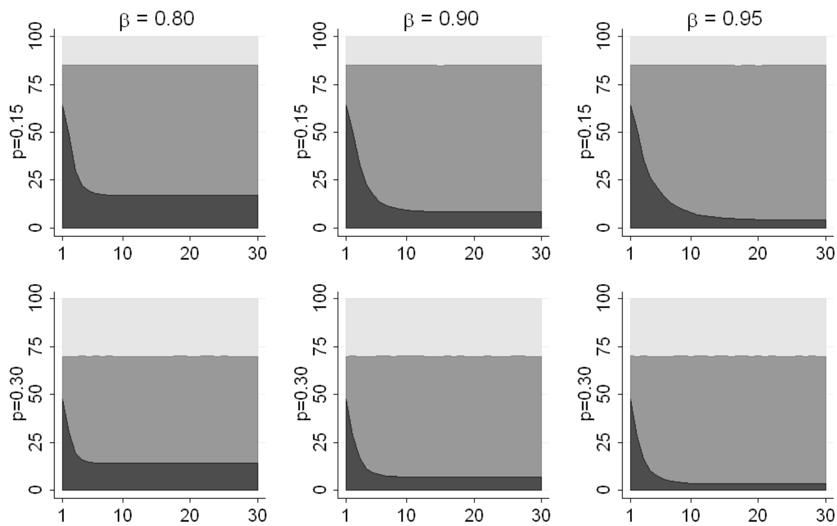
Technically, myopic best response applies to all our treatments, including the heterogeneous market treatments in which the last round's most successful choice cannot meaningfully be determined. Figure 2.7 shows simulation results for these treatments, again with initial conditions taken from pooled data from the first choices of a block (computed separately for large and small banks). Not surprisingly, predictions suffer from the same bias towards Y as imitation. The model predicts a rapid convergence to Y, while human subjects were able to maintain X choices to a large extent.

[12] There is only a single case in which the mechanics of this heuristic differ from imitation. This concerns the precise way in which the transition from X to Y takes place when three players chose X in the previous round. In fact, the way to Y is just delayed by one round.

2.4.3 Choose X when profitable

The failure of the previous models to predict our data can be ascribed to their high sensitivity to Y choices observed. As soon as players observe more than one Y, they switch to the inefficient equilibrium and are unlikely to get out of it again. It is noteworthy that with two Y choices, those who chose X still made a positive profit of 1, though it is no longer the best response to choose X. We modify the dynamic model in such a way that it models a player whose aspiration level is to achieve a positive payoff. The player chooses X if it yielded a positive payoff in the round before, and Y otherwise. When using initial propensities and experimentation mechanics as before, predicted X choice frequencies are still too low compared to the observations, though predictions are somewhat improved. We therefore further modify this heuristic.

Figure 2.7 **Simulation results for the myopic best-response heuristic (heterogeneous markets)**



Following the traditional approach, we assumed that experimentation takes place in a random and unbiased fashion. This means that players deviate from their default choice with the same probability in either direction. This is plausible if we interpret experimentation as either a decision error or an untargeted trial-and-error procedure. In our game this setting may appear less appropriate. Note that the game already involves frequent forced experimentation in the form of disruptions. Thus, if the heuristic prescribes playing X, a player will already ‘experiment’ Y with a considerable probability. It may seem appropriate to define different probabilities of experimentation depending on which option is chosen by the heuristic. We reformulate the previous heuristic as follows:

- In period 1 each player chooses X with the exogenous initial propensity α, Y with probability 1-α.
- In every following period t, determine whether choosing X would have yielded a positive absolute profit in period t-1.
- If yes, choose X with probability γ, Y with probability 1-γ.
- If no, choose Y with probability β, X with probability 1-β.

Figure 2.8 shows simulation results with γ=1, that is, the most extreme case in which all experimentation away from X is forced through disruptions. For the homogeneous treatments this model is the best so far to describe actual behaviour. It captures the persistence of the efficient equilibrium if the disruption probability is 15%, the quick trend towards Y choices in the 45% disruption case and predicts intermediate rates for 30% disruption probability (though it overstates the decline in X choices).

Figure 2.8 **Simulation results for the Choose-X-if-profitable heuristic with asymmetric experimentation (homogeneous markets)**

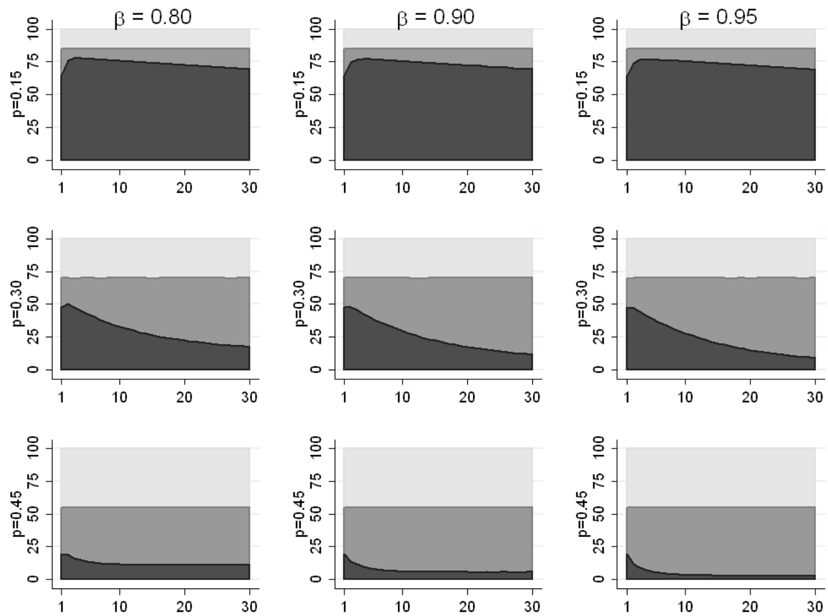
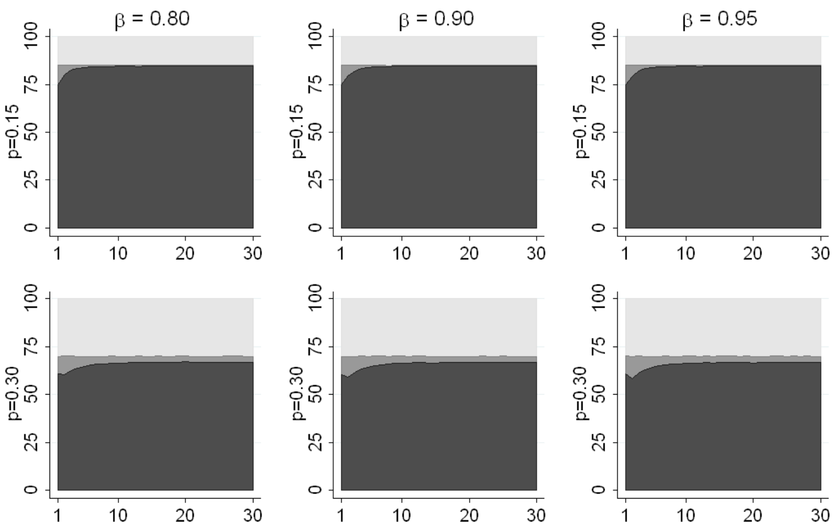


Figure 2.9 shows simulation results for the heterogeneous case. In fact, as we observe in the data, the model predicts more frequent X choices than in the homogeneous conditions. However, quantitatively the model overshoots by a long way, since it predicts a very low fraction of Y choices for 30% disruption probability. In this case the model turns out to be too tolerant towards Y choices: The large bank would choose X even if all but one of the small banks have chosen Y (since two Y choices from the small banks plus an own X choice would still leave a profit). As a result the large bank rarely switches to Y in the simulations.

In summary, none of the simple dynamics succeeds to capture all the main characteristics of all treatments of our data. Imitation and myopic best response models predict a rapid trend towards the inefficient equilibrium for all cases, which we do not observe in our data. The more tolerant heuristic to stick to the efficient equilibrium choice as long as it is profitable does considerably better, especially if it allows for experimentation to be selective. In this case the main characteristics we observe are captured. The model also qualitatively predicts that the efficient equilibrium is chosen more often in the heterogeneous market case, but it overpredicts the quantitative difference between the two types of markets.

Figure 2.9 **Simulation results for the Choose-X-if-profitable heuristic with asymmetric experimentation (heterogeneous markets)**



2.5 Conclusions

In this paper we used a stylised coordination game of Bech and Garratt (2006) to experimentally study bank behaviour in a large value payment system that is hindered by disruptions. We draw the following conclusions.

First, once behaviour moves in the direction of coordination on the inefficient equilibrium, it is not likely that behaviour moves back to the efficient equilibrium (cf. observation 2 on 'monotonicity'). The reason for this is that one player has to take the lead in going for the efficient equilibrium, but this is costly if other players do not follow suit. Analysis of different types of heuristics shows that our data is best explained by a rule of thumb in which players go for the efficient equilibrium as long as it is profitable (ie yielding a positive payoff; see section 2.4.3). In the context of a payment system, these findings suggest that once a trend has been established it is unlikely to reverse. In a situation where some banks begin to defer their payments, an intervention from the central bank may be highly desired. When banks do not have access to sufficient liquidity – ie they are forced to go for the inefficient equilibrium – central banks can use their discount window to relieve market stress. If some (critical) banks deliberately delay payments without having liquidity problems, the central bank can use its authority to encourage banks to start paying earlier (cf. Chaudhuri et al. (2009) who study the role of advice in coordination games). Such moral suasion only works though if the payment system has not been disturbed totally (ie coordinated fully on the inefficient equilibrium). So, once coordination failures start emerging central banks need to react quickly, otherwise trust between banks might have fully vanished and coordination on the 'good' equilibrium becomes highly unlikely. Note that in our experimental study there was no role for the central bank. We believe that extending the game by allowing central bank interventions would be an interesting avenue for future experimental work.

Second, coordination on the efficient equilibrium turns out to be easier in a heterogeneous market where there is clear leader in terms of size. If such a leader goes for the efficient equilibrium, 90% of smaller players who have a choice follow the leader. If the leader is not cooperative for several rounds in a row (forcedly or deliberately), the smaller players rapidly move to such a strategy as well. Given the critical role of the large player for the system as a whole, it is essential from a payment system perspective to minimise the chance that large banks are not able to execute payments due to own technical problems. It may therefore be desirable to oblige such critical participants to take extra safety measures with regard to their technical infrastructure.

Finally, our experiment shows that small frictions in coordination games can be absorbed easily and need not jeopardize the stability of the efficient equilibrium (cf. the 15% disruption cases). However, when friction becomes larger, the system can move quickly to the undesired equilibrium and stays there. In the context of payment systems this suggests that it is very important to closely monitor the payment flows of (critical) participants in the system. If deviant payment behaviour is observed by one or more participants it is important to find the reason for this behaviour. If the cause is a technical problem of one participant, the other participants in the payment system should be informed about the incident. In this way it may be avoided that the other participants falsely conclude that the deviant behaviour is a deliberate action, for example, related to liquidity considerations. Such communication is especially important during times of market stress, when false rumours can easily arise. Since we did not study communication in our experiment, it is an open research question whether this could work or not.

Appendix

Instructions of the homogeneous market case

The instructions for the homogeneous market case are shown below. Between the different experimental treatments only the percentages change. The instructions listed here are for the 15%−30%−15% case.

Instructions

Welcome to this experiment. The experiment consists of three parts in which you will have to make decisions. In each part it is possible to earn money. How much you earn depends on your own decisions and on the decisions of other participants in the experiment. At the end of the experiment a show-up fee of 5 euros plus your total earnings during the experiment will be paid to you in cash. Payments are confidential, we will not inform any of the other participants. In the experiment, all earnings will be expressed in Talers, which will be converted in euros according to the exchange rate:

1 Taler = 6 Eurocents.

During the experiment you will participate in a group of 5 players. You will be matched with the same players throughout the experiment. These other players in your group will be labeled: P2, P3, P4, and P5. You will not be informed of who the other players are, nor will they be informed of your identity.

It is not permitted to talk or communicate with others during the experiment. If you have a question, please raise your hand and we will come to your desk to answer it.

Warning: In this experiment you can avoid making any loss (negative earnings). However, note that in case you end up with a loss, it will be charged against your show-up fee.

We start now with the instructions for Part 1, which have been distributed also on paper. The instructions for the other two parts will be given when they start.

Instructions Part 1
This part consists of 30 rounds. In each round you and the other four players in your group will have to choose one of two options: X or Y.

Your earnings in a round depend on your choice and on the choices of the other four players, in the following manner:

- if you choose Y your earnings are 2 Talers regardless of the choices of the others;
- if you choose X your earnings depend on how many of the other players choose Y.

Your exact earnings in Talers from choosing X or Y, for a given number of other players choosing Y, are listed in the following table. This earnings table is the same for all players.

Number of other players choosing Y	Your earnings from choosing X	Your earnings from choosing Y
0	5	2
1	3	2
2	1	2
3	-1	2
4	-3	2

For example, if 2 other players choose Y, then your earnings from choosing X will be 1, while your earnings from choosing Y would be 2.

Forced Y

Note, however, that you may not be free to choose your preferred option. In each round, each of you will face a chance of 15% that you are forced to choose option Y. We will call this a 'forced Y'.

Whether or not a player is forced to choose Y is randomly determined by the computer for each player separately and independently from the other players. Further, a forced Y does not depend on what happened in previous rounds.

On the computer screen where you take your decision you will be reminded of this chance of a forced Y, for your convenience. Furthermore, in the table at the bottom of that screen (showing past decisions and earnings) your forced Y's are indicated in the column showing your choices with an 'F'. Note that you will not be informed of other players' forced Y choices.

You are now kindly requested to do a few exercises on the computer to make you fully familiar with the earnings table. In these exercises you cannot earn any money.

Thereafter, we will start with Part 1.
Please raise your hand if you have any question. We will then come over to your table to answer your question.

Instructions Part 2
Part 2 is exactly the same as Part 1, except for one modification.
In each round, each of you will now face a chance of 30% that you are forced to choose option Y.
Are there any questions?

Instructions Part 3
Part 3 is exactly the same as Part 2, except for one modification.
In each round, each of you will now face a chance of 15% that you are forced to choose option Y, like in Part 1.
Are there any questions?

Instructions of the heterogeneous case
The instructions for the heterogeneous market case are shown below. Again, between the different experimental treatments only the percentages change. The instructions listed here are for the 15%−30%−15% case.

Instructions

Welcome to this experiment. The experiment consists of three parts in which you will have to make decisions. In each part it is possible to earn money. How much you earn depends on your own decisions and on the decisions of other participants in the experiment. At the end of the experiment a show-up fee of 5 euros plus your total earnings during the experiment will be paid to you in cash. Payments are confidential, we will not inform any of the other participants. In the experiment, all earnings will be expressed in Talers, which will be converted in euros according to the exchange rate: 1 taler = 6 euro cents.

During the experiment you will participate in a group of 4 players. You will be matched with the same players throughout the experiment. There are two types of players: A and B. The difference is related to the consequences of their decisions, as will be explained below. In fact, there will be 1 A player and 3 B players in your group. If you happen to be player A then the others are B players, who will be labeled B1, B2, and B3. If you are a B player then the other players in your group comprise a player A and two other B players, denoted as

B2 and B3. You will learn your type when Part 1 starts; it will stay the same during the whole experiment,. Because we have pre-assigned a type to each table, you have drawn your type yourself when you selected a table number in the reception room. You will not be informed of who the other players are, nor will they be informed of your identity.

It is not permitted to talk or communicate with others during the experiment. If you have a question, please raise your hand and we will come to your table to answer it.

We start now with the instructions for Part 1, which have been distributed also on paper. The instructions for the other two parts will be given when they start.

Instructions Part 1

First of all, note that your type (A or B) will be shown at the upper-left part of your computer screen, below a window showing the round number.

This part consists of 30 rounds. In each round you and the other three players in your group will have to choose one of two options: X or Y. Your earnings in a round depend on your type (A or B), your choice, and the choices of the other three players, in the following manner:

- if you choose Y your earnings are 2, regardless of your type and the choices of the others;
- if you choose X your earnings depend on your type and on how many of the other players choose Y.

Your exact earnings from choosing X or Y, given your type and the Y choices of the other players in your group, are listed in the following tables for, respectively, player A and a B player.

Some examples, for illustration.

Suppose you are a player A, and you choose X while 1 of the other players chooses Y, then the upper table shows that your earnings will be 3.

Alternatively, suppose you are a B player, and you choose X while 1 of the other players chooses Y, then it depends on whether this other player choosing Y is a player A or another B player. If it is player A, then the lower table shows that your earnings are 1, while your earnings are 3 if it is a B player. Thus, player A has a larger impact on your earnings than a B player.

Player A

Your choice	Number of B players choosing Y	Your earnings
X	0	5
X	1	3
X	2	1
X	3	-1
Y	0	2
Y	1	2
Y	2	2
Y	3	2

Player B

Player A's choice	Number of other B players choosing Y	Your earnings from choosing X	Your earnings from choosing Y
X	0	5	2
X	1	3	2
X	2	1	2
Y	0	1	2
Y	1	-1	2
Y	2	-3	2

Forced Y

Note, however, that you may not be free to choose your preferred option. In each round, each of you will face a chance of 15% that you are forced to choose option Y. We will call this a 'forced Y'.

Whether or not a player is forced to choose Y is randomly determined by the computer for each player separately and independently from the other players. Further, a forced Y does not depend on what happened in previous rounds.

On the computer screen where you take your decision you will be reminded of this chance of a forced Y, for your convenience. Furthermore, in the table at the bottom of that screen (showing past decisions and earnings) your forced Y's are indicated in the column showing your choices with an 'F'. Note that you will not be informed of other players' forced Y choices.

You are now kindly requested to do a few exercises on the computer to make you fully familiar with the earnings table. In these exercises you cannot earn any money.

Thereafter, we will start with Part 1.
Please raise your hand if you have any question. We will then come over to your table to answer your question.

Instructions Part 2

Part 2 is exactly the same as Part 1, except for one modification.
In each round, each of you will now face a chance of 30% that you are forced to choose option Y.
Are there any questions?

Instructions Part 3

Part 3 is exactly the same as Part 2, except for one modification.
In each round, each of you will now face a chance of 15% that you are forced to choose option Y, like in Part 1.
Are there any questions?

References

Abbink, K − Brandts, J (2008) **Pricing in Bertrand competition with increasing marginal costs.** Games and Economic Behavior, 63(1), 1−31.

Alós-Ferrer, C − Ania, A − Schenk-Hoppé, K (2000) **An evolutionary model of Bertrand oligopoly.** Games and Economic Behavior, 33, 1, 1−19.

Bech, M − Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory 109, 198−210.

Bech, M − 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 − Hobijn, B (2007) **Technology diffusion within central banking: the case of real-time gross settlement.** International Journal of Central Banking, 3, 147−181.

Bech, M (2008) **Intraday liquidity management: a tale of games banks play.** FRBNY Economic policy review.

Chaudhuri, A − Schotter, A − Sopher, B (2009) **Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice.** Economic Journal, 119, 91−122.

Crawford, V (1995) **Adaptive dynamics in coordination games.** Econometrica, 63(1), 103−143.

Cubitt, R P − Sugden, R (1999) **The selection of preferences through imitation.** Review of Economic Studies, 65, 761−771.

Devetag, G − Ortmann, A (2007) **When and why? A critical survey on coordination failure in the laboratory.** Experimental Economics, 10, 311−344.

Friskies Gourmet News (2003) **Katzen ahmen unsere Sprache nach – Interview mit Nobelpreisträger Univ.-Prof. Dr. Reinhard Selten.** Issue 9 (June), 2.

Harsanyi, J − Selten, R (1988) **A general theory of equilibrium selection in games.** MIT Press.

Heijmans R H (2009) **Simulation in the Dutch interbank payment system: a sensitivity analysis.** Proceedings from the Bank of Finland Payment and Settlement System Seminars 2007−2008, 124−144.

Kahn, C − McAndrews, J − Roberds, W (2003) **Settlement risk under gross and net settlement.** Journal Money and Banking 35, 591−608.

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

Ledrut, E (2008) **Simulating retaliation in payment systems: can banks control their exposure to a failing participant.** DNB working paper 133.

McAndrews, J − Potter, S (2002) **Liquidity effects of the events of September 11th.** Federal Reserve Bank of New York Economic Policy Review 8, 59−79.

McAndrews, J − Rajan, S (2000) **The timing and funding of Fedwire funds transfers.** Federal Reserve Bank of New York Economic Policy Review 6, 17−32.

Mehta, J − Starmer, C − Sugden, R (1994) **The nature of salience: an experimental investigation of pure coordination games.** American Economic Review 84, 658−673.

Pröpper, M − van Lelyveld, I − Heijmans, R (2008) **Towards a Network Descriptions of Interbank Payment Flows.** DNB working paper 177.

Schelling, T C (1960) **The strategy of conflict.** Harvard University Press.

Schlag, K H (1998) **Why imitate, and if so, how? A Boundedly Rational Approach to Multi-armed Bandits.** Journal of Economic Theory, 78, 130−156.

Selten, R − Ostmann, A (2001) **Imitation equilibrium.** BonnEcon Discussion Paper 16/2001, University of Bonn.

Soramäki, K − Bech, M − Arnold, J (2007) **The topology of interbank payment flows.** Physica A, 379(1), 317−333.

Van Huyck, J − Battalio, R − Beil, R (1990) **Tacit coordination games, strategic uncertainty and coordination failure.** American Economic Review, 80, 234−248.

Van Huyck, J − Battalio, R − Beil, R (1991) **Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games.** Quarterly Journal of Economics, 106, 885−910.

Vega-Redondo, F (1997) **The evolution of walrasian behavior.** Econometrica, 65, 375−384.

Chapter 3

Methods for evaluating liquidity provision in real-time gross settlement payment systems*

Edward Denbee − Rodney Garratt – Peter Zimmerman

** The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. The authors wish to thank Evangelos Benos, Paul Chilcott, Roger Clews, Toby Davies, Marco Galbiati, Matti Hellqvist, Matthew Hunt, John Jackson, Edwin Schooling Latter, Anne Wetherilt, an anonymous referee and seminar participants at the Bank of England and at the Bank of Finland 8th Simulator Seminar and Workshop 2010 for useful comments and suggestions, and Lewis Webber for invaluable help with simulations. All errors are ours.

3 Methods for evaluating liquidity provision in real-time gross settlement payment systems

Abstract

Banks must provide costly intraday liquidity in order to settle payments in real-time gross settlement systems. They can provide this liquidity themselves, or they can wait for incoming payments to fund subsequent outgoing payments. However, if too many banks withhold liquidity in this way then the result could be gridlock in the payments system. We develop new methods to detect this behaviour empirically, examining whether banks supply an amount of liquidity to the system commensurate with the share of payments they are responsible for. In aggregate, it appears that in the period after the collapse of Lehman Brothers there may have been increased mismatches between liquidity usage and liquidity provision in CHAPS, the UK large-value payment system. We also model payments activity as a random walk and use a recombinant approach coupled with a quantile regression technique to detect apparent free-riding behaviour. We present these measures using data from CHAPS as an example.

3.1 Introduction

A payment system consists of the procedures and associated computer networks used by both domestic and international financial institutions to transfer funds, securities and derivatives. Sometimes called the 'plumbing' of the financial system, smoothly functioning payments systems are essential to the operation of financial markets. Large value wholesale payment systems, such as CHAPS in the United Kingdom, are particularly important because of their tremendous size and the importance of the financial transactions they facilitate. On a typical business day, transactions valued at nearly £250 billion flow through CHAPS, roughly equivalent to one-sixth of the United Kingdom's annual gross domestic product.[1]

[1] See Bank of England (2011).

In a real-time gross settlement (RTGS) system such as CHAPS, processed payments settle immediately and with complete finality. Thus there is no credit risk from timing mismatches or delays once payments have been submitted to the system. However, real-time gross settlement means that the paying bank has to be able to fund the payment upfront: banks must have liquidity available in their central bank accounts to fund any outgoing payments.[2]

To facilitate payment activity, settlement banks use their holdings of cash reserves along with additional liquidity that they obtain via repos of eligible securities (usually very high-quality industrialised country sovereign bonds) with the central bank.[3] However, banks do not have to provide enough liquidity to fund their entire gross payment flows. During the course of the day banks make and receive thousands of payments. Thus banks use their own liquidity to make payments in combination with liquidity obtained from incoming payments.

Liquidity in the payment system is therefore a public good in the sense that once some banks provide it to make payments, direct or indirect recipients can reuse it – over and over again – to make their own payments. If banks were required to process payments as soon as they receive them then liquidity provision of each individual bank would be largely exogenous. But this is not the case: with the exception of some time-critical payments, banks do not have to process payment requests as soon as they receive them. Rather, banks may choose to delay processing payments in order to conserve liquidity and make use of incoming funds.[4]

If too many banks withhold liquidity the payment system can fall into gridlock.[5] Consequently, central banks may have an interest in monitoring banks' liquidity provision in order to ensure the continued smooth functioning of the payments system. A bank may appear to

---

[2] In RTGS systems equipped with liquidity saving mechanisms, this may no longer be true. See Norman (2010).

[3] The term settlement bank refers to a bank that has a settlement account with the central bank for the purposes of making intraday payments through the payments system. There are currently 18 settlement banks in CHAPS. Other banks that operate in the United Kingdom make CHAPS payments via one or more of these settlement banks.

[4] The incentive to conserve liquidity arises because funds that banks deposit in their settlement accounts to facilitate payments have an opportunity cost to the bank in terms of foregone investment opportunities, or to mitigate against the possibility of liquidity shocks later in the day.

[5] CPSS (1993) defines a gridlock as a situation where 'the failure of some transfer instructions to be executed...prevents a substantial number of other instructions from other participants from being executed'. See Soramäki and Bech (2001) for a discussion of the cost of delay in payment systems.

free-ride when the share of liquidity that it provides to the payments system is less than the share of liquidity that it uses. We use the term 'apparent free-riding', because we are unable to judge whether the bank is deliberately conserving its liquidity,[6] or whether this difference between liquidity usage and provision is an unintended consequence of its business model.

This paper presents new measures developed to detect apparent free-riding in RTGS payments systems. We begin by looking at free-riding across two dimensions, cost and risk. The cost measure is related to the amount of liquidity that a bank provides whereas the risk measure relates to how prepared a bank is to let its liquidity circulate in the system, exposing it to counterparty risks.[7]

Our measures identify apparent free-riding in a strict accounting sense; that is, when a bank takes a smaller share of the liquidity cost or risk than its share of gross payment values it is a free-rider. However, these measures do not distinguish between free-riding that results from intentional delay and that which occurs for structural reasons. For instance, a bank which sends and receives a small number of very large payments would be expected to use more liquidity than a bank with a large number of smaller payments. This is because the latter bank would be better able to net off payments and receipts and thus less likely to assume a large net sender position.[8] Also, banks may not be in a position to make payments early in the day. For example, consider the case of a CHAPS payments bank with a lot of clients based in the US. These clients might only send same-day instructions in the UK afternoon, due to time zone differences. Such a bank may appear to be hoarding liquidity when in fact it may be making its payments as soon as the instructions arrive. Conversely, some payments are intraday time-critical – that is, they must be made by a particular time of day. This means that the bank may appear to be generous with its liquidity when in fact the decision to pay early is driven by exogenous factors.[9]

We compute cost and risk based free-riding measures for CHAPS settlement banks and present aggregated results for groups of banks in

---

[6] See Norman (2010) or Ball et al (2011).

[7] Having paid away part of its liquidity, if counterparties are slow in sending back to a bank, then the bank has to support its remaining payments on limited liquidity. And, if a counterparty defaults during the day the bank may have to claim its incoming payments from an administrator, which could be a lengthy process. It may be much easier simply to hold back payments and net out obligations with the defaulted estate.

[8] This is the liquidity pooling effect – see Jackson and Manning (2007).

[9] According to Ball et al (2011), such time-critical payments comprise only around 4% of values in the system.

two size categories. We find that banks with higher daily payment values appear to free-ride on cost more than banks with lower values. The relationship between size of payment flows and apparent free-riding is similar in the case of liquidity risk. The banks with greater flows appear to free-ride off the others, while those with smaller flows take on more of the intraday liquidity risk. However, the banks with greater flows tend to have different business models to those with lower flows – for example, they tend to have more wholesale than retail payments, and many clear payments on behalf of other institutions – and so it is not possible to conclude whether the differences in apparent free-riding behaviour are driven by size alone.

By summing up the amount of apparent free-riding across banks we obtain measures of the aggregate amount of free-riding in the system. We analyse these measures over time to see how the amount of free-riding varied during the period of financial crisis following the collapse of Lehman Brothers. We find that aggregate free-riding of both types rose during this period.

It would be desirable to identify free-riding that occurs because a bank strategically delays payments in order to benefit from the liquidity provision of other banks.[10] We provide a method for identifying intentional free-riders by controlling for some of the differences in banks' payment characteristics.

We account for differences in the timing of banks' payment request arrivals by recombining each bank's historic payments in order to simulate a payments day. By randomly re-ordering the payments, we produce a simulated day where all the same payments are made. That is, total amounts sent and received between each pair of banks is the same as in reality — but there is no behavioural bias in terms of the order of payments, and therefore liquidity usage. Based on this model, we use quantile regression to find a threshold for a bank's free-riding measures, beyond which we can say there is significant evidence that the share of liquidity that the bank provides to the system is smaller than the share it should be expected to provide, based on its usage.

[10] We emphasise intentional delay based on liquidity concerns. However, it is also possible that settlement banks, when making payments on behalf of clients, intentionally delay in order to reduce credit exposures to these clients. According to Valukas (2010), Lehman Brothers' settlement banks tried to reduce their unsecured intraday exposures to the institution once its financial condition began to deteriorate.

3.2 Empirical measures of free-riding

3.2.1 Cost-based measure

One way to measure free-riding is to look at the share of total liquidity a bank provides to the system and to relate this to its share of total payments. Banks provide liquidity to the system when they send more than they have received. The difference has to be made up either from central bank reserves, or from eligible collateral that settlement banks repo intraday in order to obtain liquidity from the central bank.[11] Therefore we can assume that a net debit position imposes a cost upon the bank – this is the opportunity cost of using central bank reserves or of pledging eligible collateral.

Suppose there are n banks, which are indexed by i=1,...,n. Let $x_i^s(t)$ be the amount sent by bank i up to time t on day s, and let $y_i^s(t)$ be the amount received. t lies in the interval [0,T], where t=0 denotes the start of the day and t=T at the end. Then the net debit position at time t on day s is

$$N_i^s(t) = x_i^s(t) - y_i^s(t)$$

The net debit position identifies the liquidity provided by bank i to the rest of the system by time t on day s. Therefore the liquidity burden of bank i on day s is determined by the largest net debit position

$$L_i^s = \max_{t \in [0,T]} N_i^s(t)$$

Note that $L_i^s \geq 0$, since $x_i^s(0) = y_i^s(0) = 0$ for all i, s.

The largest net debit position incurred by a bank on a given day is the total amount of the bank's own cash and collateral that it actually used to fund its own payments. It is the minimum amount of liquidity that the bank could have held to meet its payment obligations on that day given the behaviour of others, and so reflects the cost burden. This

[11] Banks can also fund payments using liquidity received in interconnected payments systems. For instance, banks which are settlement banks in both CHAPS and CREST – the UK securities settlement system – could use a net receiver position in one system to fund payments in the other. We do not consider this in our analysis, partly because in the UK these other interconnected payments systems do not allow settlement banks the same degree of control over intraday timing as CHAPS does.

may be the (opportunity) cost of acquiring additional central bank reserves or eligible collateral.

A bank is an apparent free-rider if it uses a larger share of system liquidity than the share it provides. This is valid in a strict accounting sense because if a bank makes a larger share of payments than the share of liquidity it provides, then it is free-riding on the liquidity of the banks that provided a greater share.

The cost-based ex-post measure of free-riding for bank i on day s is defined as

$$c_i^s = \frac{L_i^s}{\sum_j^n L_{j=1}^s} - \frac{x_i^s(T)}{\sum_{j=1}^n x_j^s(T)}.$$

This measures the difference between bank i's share of liquidity provision and its share of liquidity usage. A value close to –1 would mean that the bank made almost all the payments in the system that day but provided almost no liquidity. At the other extreme, a value close to 1 would mean that a single bank provided almost all the liquidity but made a small proportion of the payments.[12] A bank that provided the same share of liquidity as its share of total payments would have a value of 0. Hence we can say a bank appears to be a free-rider if it has a value less than 0 and appears not to free-ride if it has a value greater than 0.

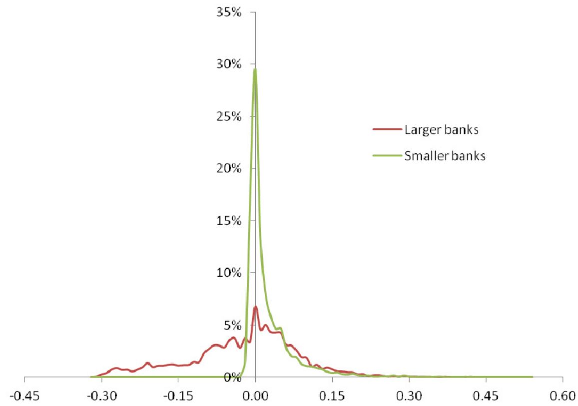
Chart 3.1 is a histogram showing the observed free-riding on cost measures over the period 1 January 2008 to 31 May 2010.[13] We split the banks into two groups according to average daily values sent over the period: the larger banks are coloured red while the smaller banks are green. Larger banks account for over 90% of total payment values sent through CHAPS over this period. The makeup of each group is recalibrated on 5 February 2009.[14] We emphasise that 'larger' and 'smaller' here refer only to average daily values through CHAPS over the period and are not necessarily correlated with other measures of size, such as balance sheet size or payments in non-sterling currencies.

[12] It cannot ever actually attain the values –1 or 1, because a bank which provides any positive amount ε of liquidity must by definition send at least ε of payment values that day.

[13] The bin size on the horizontal axis is 0.01, with peaks occurring in the interval [-0.005, 0.005].

[14] A merger between two CHAPS settlement banks led to one bank's reserves account being withdrawn on 5 February 2009, meaning that those two settlement banks effectively operate from a single pool of liquidity after this date.

Chart 3.1 **Frequencies of observed values of c_i^s for larger and smaller banks over the period 1 January 2008 to 31 May 2010**



The chart shows that the larger banks tend to have the widest range of free-riding values. Over this period, the measure for smaller banks takes an average value of 0.02 while for larger banks it is –0.02. Difference of means tests reveal significance: the banks classified as larger appear to free-ride on cost more often than those classified as smaller. Of course there is no evidence to suggest this is strategic behaviour. It may be that larger banks structurally pay later than the than the smaller ones, due to the nature of their business. The activity of many of the banks in the 'smaller' group is dominated by retail business, while the payment activity of many of the 'larger' banks may be driven by wholesale business (for example, lending and borrowing in the money market). And many of the larger banks make payments on behalf of other institutions. These factors may lead to structural differences in banks' payment schedules.

Alternatively these results could suggest that there are liquidity efficiency benefits to being a larger bank. We explore this possibility in Section 3.3.

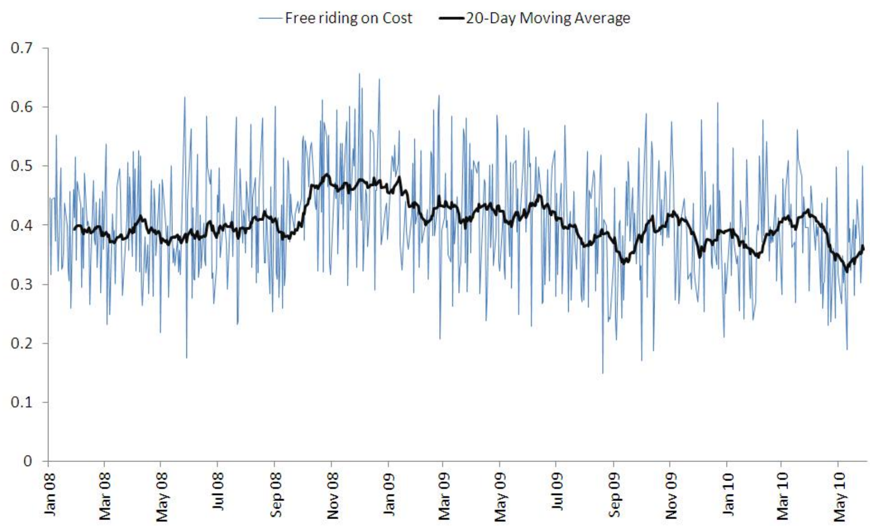
For a given s, the measure c_i^s sums over i to zero. This makes sense because banks that free-ride must do so on someone else in the system. Looking at the sum over just the free-riding banks (ie those where $c_i^s < 0$) we obtain an aggregate measure of the amount of free-riding in the system on any given day.

Define

$$m^s = \sum_{\{i:c_i^s<0\}} \left|c_i^s\right|$$

m^s has a minimum value of 0 when all banks provide their fair share of liquidity (ie when $c_i^s = 0$ for all i), and is bounded above by 1. It achieves higher values when some banks provide the bulk of the liquidity while making few payments (and equivalently, others make a greater proportion of payments than the share of liquidity they provide – that is, there is apparent free-riding in the system). Chart 3.2 plots the 20-day moving average value of m^s using data from 1 January 2008 to 31 May 2010.

Chart 3.2 **Aggregate free-riding on cost m^s in CHAPS over the period 1 January 2008 to 31 May 2010. The black line denotes the 20-day backward-looking moving average**



The chart suggests the presence of a structural break in the series in September 2008. Applying a Chow structural break test to the individual daily data over the 179 observations before and after this day, we reject at the 1% level the null hypothesis that there is no structural break on 15 September 2008. Table 3.1 below shows the results.[15] We conclude that there was a significant increase in free-riding on cost in CHAPS after the Lehman Brothers default on 15 September 2008.[16]

Table 3.1 **Chow structural break test for the free-riding on cost measure at 15 September 2008. The period tested is 1 January 2008 to 6 February 2009, which is 179 days before and after the date tested**

F-statistic	32.31
Degrees of freedom	356
p-value	0.000

Measuring free-riding based on largest net debit position has its drawbacks. While this measure addresses the direct cost to a bank of providing liquidity, it does not reflect exposure to counterparty risk. When a bank allows itself to maintain a debit position it is exposed to risk if a counterparty defaults or suffers operational problems, because then that counterparty will not be able to deliver on its obligations and so the liquidity sent will not be recycled back into the system. To avoid this, the bank may prefer to withhold its payments so that they can be offset against any obligations due (see Manning, Nier and Schanz, 2009). In a sense, by taking a net debit position the bank is, consciously or not, sacrificing itself for the common good; it is prepared to take some intraday liquidity risk to grease the wheels of the payment system. A measure attuned to counterparty risk should take into account the length of time for which the bank is exposed to its counterparties. An alternative measure of free-riding may therefore consider how unwilling the bank is to hold net debit positions for extended periods of time.

[15] We could take a longer post-Lehman sample but we run the risk that behaviour changes again once the suspected break event is distant enough. Therefore we restrict to 179 days, which is the same as the pre-Lehman sample period.

[16] Benos, Garratt and Zimmerman (2012) investigate changes in settlement bank behaviour in CHAPS following the collapse of Lehman Brothers, and suggests reasons for these changes in behaviour.

3.2.2 Risk-based measure

Rather than looking only at the largest net debit position incurred during the day, we now examine the time-weighted exposure of banks' net sender positions. This represents the share of the system-wide counterparty risk that a bank incurs, as described above. When a bank has a negative net position it sends payments more slowly than it receives them and thus, deliberately or not, hoards the liquidity that other banks have provided to the payment system. Since this bank uses a greater proportion of system-wide liquidity than the share of counterparty risk that it assumes, this can be interpreted as apparent free-riding.

A bank takes counterparty risk when it is a net sender – that is, $N_i^s(t) > 0$. The average risk taken for bank i on day s is[17]

$$\Lambda_i^s = \frac{1}{T}\sum_{t=0}^{T} \max\left[N_i^s(t), 0\right]$$

The free-riding on risk measure considers the share of the total risk each bank is prepared to take. Again, this should be compared to the share of payments a bank makes. The risk-based measure of free-riding is thus

$$\gamma_i^s = \frac{\Lambda_i^s}{\sum_{j=1}^{n} \Lambda_j^s} - \frac{x_i^s(T)}{\sum_{j=1}^{n} x_j^s(T)}$$

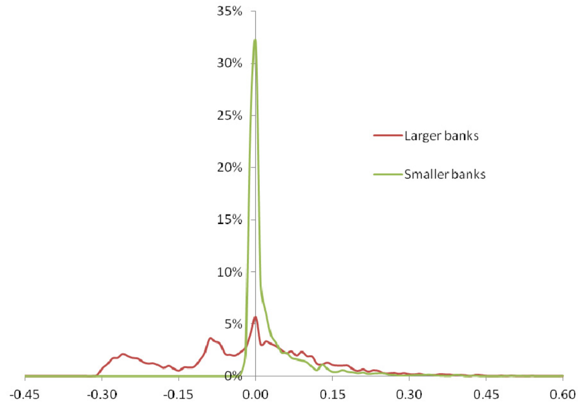
When $\gamma_i^s > 0$, bank i takes a share of the aggregate counterparty risk in excess of its payments share. When $\gamma_i^s < 0$, bank i takes on less risk than its share of payments. As with c_i^s, for a given s this measure sums to zero across all banks.

Chart 3.3 is a histogram showing observed values of the free-riding on risk measure over the sample period from 1 January 2008 to 31 May 2010. The bin size on the horizontal axis is 0.01. The banks are grouped in the same way as for the free-riding on cost measure. The relationship between size and free-riding is similar to before. The

[17] Our free-riding on risk measure can also be thought of as a measure of liquidity cost In a regime where collateral has some intraday opportunity cost – that is, where collateral can be used for other purposes intraday. Leinonen and Soramäki (1999) discusses flexible liquidity regimes like this.

larger banks appear to free-ride off the others. The larger banks have a mean free-riding on risk value of –0.02 and the smaller banks 0.03. Again difference in means tests reveal significance.

Chart 3.3 **Frequencies of observed values of γ_i^s for larger and smaller banks over the period 1 January 2008 to 31 May 2010**

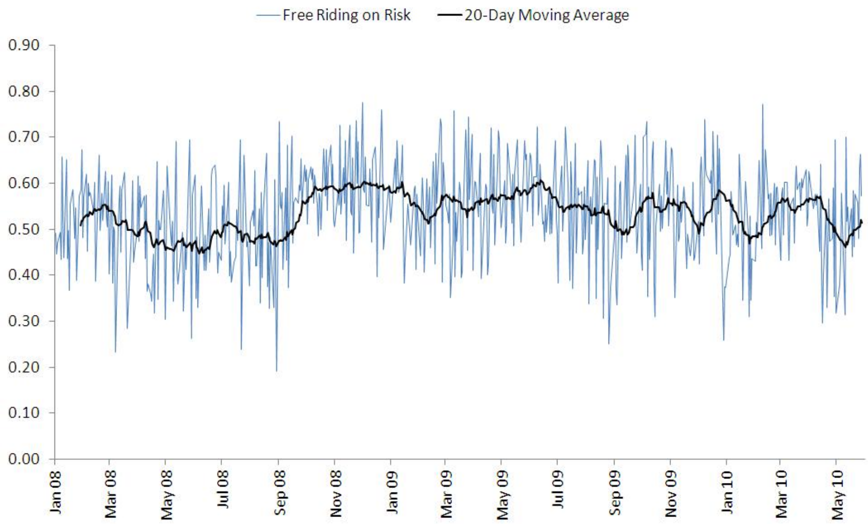


Let

$$\mu^s = \sum_{\{i:\gamma_i^s < 0\}} \left|\gamma_i^s\right|$$

This is an aggregate free-riding on risk measure, analogous to m^s. Chart 3.4 shows m^s over the period 1 January 2008 to 31 May 2010.

Chart 3.4 **Aggregate free-riding on risk μ^s in CHAPS over the period 1 January 2008 to 31 May 2010, 20-day backward-looking moving average**



As with the free-riding on cost measure, there appears to be a significant increase after the Lehman Brothers default on 15 September 2008. Again a Chow structural break test applied to the 179 observations before and after the Lehman Brothers default rejects the hypothesis that there was no structural break on this day – see Table 3.2.

Table 3.2 **Chow structural break test for the free-riding on risk measure at 15 September 2008. The period tested is 1 January 2008 to 6 February 2009, which is 179 days before and after the date tested**

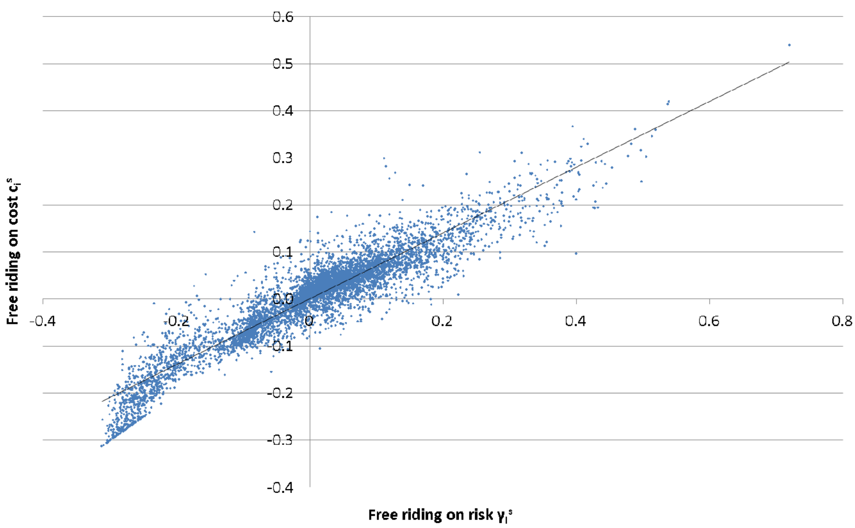
F-statistic	64.60
Degrees of freedom	356
p-value	0.000

Ball et al (2011) explain that many banks have internal schedulers which allow them to place bilateral limits against individual counterparties. By exercising these limits, a bank can control the amount of liquidity it contributes to the system, and thus indirectly manage its free-riding on cost measure. A bank that uses these limits to control the duration of its liquidity exposures would indirectly be able to manage its free-riding on risk measure. The results of this section suggest that CHAPS settlement banks aim to manage the duration as well as the magnitude of these exposures.

3.2.3 Comparing free-riding measures

These two forms of free-riding, on cost and on risk, are likely to be correlated. In most cases a bank that builds up a large net debit position will do so gradually over the course of the day. This means it takes a share of the liquidity risk associated with having sent funds to other banks for extended periods of time. The correlation is shown in Chart 3.5, which plots the paired free-riding measures for each of the settlement banks on each day over a period extending from 1 January 2008 to 31 May 2010. Banks that free-ride in both dimensions appear in the top right corner of the chart, whereas banks that are providers in both directions appear in the bottom left. This correlation is positive and significant, with an adjusted R^2 value of 87.6%.

Chart 3.5 **The relationship between free-riding on cost and free-riding on risk, 1 January 2008 to 31 May 2010. The black line represents the best fit to a linear relationship**



3.3 Random walk models of free-riding

3.3.1 Modelling payments processes as random walks

The metrics introduced in Section 3.2 assume that banks' contributions to liquidity provision (as measured on a cost or risk basis) should be proportionate to their share of liquidity usage. In this Section we test that assumption. We also find critical values for the free-riding measures, such that if these values are exceeded then we can say that there is significant evidence that the bank is an apparent free-rider.

The need for a bank to make payments during the day comes from activity in the real economy and its own interbank obligations. We can model each bank's net debit position as a random walk process – that is, payments arise according to some distribution and the position of each sending and receiving bank moves up or down accordingly. In the absence of any information about what factors may be driving the timing of the arrival of payments, a random walk provides us with a way of modelling the evolution of the net debit position, requiring only assumptions about the distribution of payment values.

An interesting feature of random walks is that their expected maximum or minimum positions do not vary linearly with the length of the walk (which corresponds to the number of payments). For example, a bank that has twice as many incoming and outgoing payments as another bank may require less than twice the liquidity to cover its largest net debit position.[18] This means that our measures as presented in the previous section may unfairly label larger banks as free-riders. For example, as a bank increases its share of values sent, its share of liquidity provision is likely to increase by a smaller proportion, thus reducing its free-riding on cost measure c_i^s. Based on these arguments, we might argue that a bank should only be classed as a free-rider if it provides less liquidity or takes on less risk than expected, *given* the schedule of payments that it is due to make.

For these reasons we compute a measure of expected free-riding that attempts to account for differences in banks' payments schedules. This will be the benchmark against which banks' actual behaviour is compared. We compute a threshold value for free-riding for a generic

---

[18] Galbiati and Giansante (2010) model liquidity usage as a symmetric random walk. In their model, it can be shown that expected liquidity usage varies asymptotically with the square root of payment volumes.

bank, under the assumption that all payments are processed immediately and that requests for payments in and out of a bank arrive according to a random process. This means that if the bank's observed measure falls below the threshold we say it appears to be a free-rider. We wish to select the threshold so that there is a 5% probability of a Type I error. To do this we use quantile regression, as developed by Koenker and Bassett (1978).

3.3.2 Random walk using historical payment distributions: a recombinant approach

In this subsection, we present a model which allows us to identify apparent free-riding behaviour while controlling for banks' size. We assume that the payments that must be made on a given day are exogenous, but that their ordering is subject to individual banks' behaviour. Therefore each bank can be considered to have a 'file' of payments that it needs to send on a particular day, but the actual timing of each payment does not matter, so long as it is made before the end of the day. If we randomly re-order the payments, we can produce a simulated day where all payments are made – that is, total amounts sent and received between each pair of banks is the same as in reality – but the actual liquidity used by each bank may be quite different, because liquidity usage depends to a large extent on the order that payments are made and received. This allows us to factor out behavioural biases, where in order to reduce liquidity usage the bank may elect to wait to receive before it sends.

By simulating a large number of random re-orderings of the payments file, we can calculate the amount of liquidity that the bank might use to make these payments absent of behavioural biases. Then, if the actual amount of liquidity used is significantly less than this, the bank can be labelled as an apparent free-rider.

There may still be structural reasons why payments need to be made in a certain order during the day, as well as behavioural aspects. For example, there may be payments which are made on behalf of customers which only send their payment orders through in the afternoon. We are unable to identify these, and thus cannot correct for them.

3.3.2.1 Simulation

Since the distribution of payments on each day varies, we need to consider a period of several days. We use 102 days of data (1 January 2010 to 31 May 2010). Each day's simulation involves randomly reordering around 125,000 transactions and so it is too computationally expensive to consider a longer period. Moreover, if we take a longer period there is the risk of structural changes in banks' liquidity usage (for example, when a settlement bank takes on a large new customer).

Further it is not feasible to consider all possible permutations for every day. Instead we simulate each day 200 times. We find that 200 is sufficient to produce stable empirical results.

We treat payments from or to the Bank of England and CLS Bank as exogenous, since these settlement banks do not have incentives to engage in strategic free-riding on liquidity. These payments therefore retain their order in the simulations. For example, if such a payment was the tenth payment of the day in reality, then it remains the tenth payment in our recombinant simulations of that day.

3.3.2.2 Regression

We use quantile regression to estimate the 5% confidence threshold for apparent free-riding. The free-riding measures obtained from the simulations are regressed against individual bank characteristics. This is equivalent to solving the following minimisation problem

$$\hat{\beta} = \arg \min_{\beta \in \Re^k} \sum_{j=1}^{N} \rho(y_j - X_j^{'}\beta)$$

where N is the number of observations, k the number of independent variables (which are bank characteristics), y_j the observed free-riding measure from simulation, X_j a k-vector of bank characteristics for observation j, and $\rho(\theta) = \theta(0.05 - I\{\theta < 0\})$ is a loss function.[19] Since

[19] This 'tick function' takes the value 0.05θ if $\theta \geq 0$ and -0.95θ if $\theta < 0$. Koenker and Bassett (1978) show that minimising this loss function will deliver a coefficient $\hat{\beta}$ which exceeds the observed data with probability 5%.

each of the 102 days is simulated 200 times and there are 12 banks, we have $N = 102 \times 200 \times 12 = 244{,}800$ observations.[20]

The vectors X_i should contain individual bank characteristics that we believe may affect the degree of free-riding. We wish to capture some measure of 'size' of the bank (in payment system terms). A bank which sends a higher value of payments will, in general, provide more liquidity to the system and appear to be less of a free-rider, while a bank with a higher value of payments received will tend to appear to free-ride more. What is important for our free-riding measures is value sent/received compared to other banks in the system, so we use values as a proportion of total system activity (that is, total value including payments made by Bank of England and CLS Bank). We find that for our data values sent and values received are highly collinear, so in our regressions we just use values sent.[21]

We may expect sizes of individual payments to influence our results too. Consider the simulation carried out on two hypothetical banks with identical total payment values, but assume that bank 1 has a larger average payment size than bank 2. When payments are randomly ordered in the simulations, bank 1 would have a higher probability of achieving a large net receiver position than bank 2.[22] Therefore we might expect a bank with a higher average payment size to appear to be more of a free-rider. This is still true even if it is only payments sent or only payments received which have a larger average size.

In addition to using total value sent, average sent payment size and average received payment size, we consider using volumes sent and volumes received. 'Volume' refers to the number of payments sent by a bank on day, unweighted by value. Again, we use volume as a proportion of total system volumes. We find that volume sent is collinear with the other regressors, but that volume received is not.

[20] Bank of England and CLS Bank have been removed from the set of dependent variable observations, since we would not expect their liquidity usage to be determined by the same factors as the other settlement banks.

[21] We use the following diagnostics for multicollinearity in this section: variance inflation factor (VIF) of less than 10; condition number of less than 15; and bivariate R^2 between all pairs of regressors less than 0.9. If all three of these conditions are satisfied, then we conclude that multicollinearity is not a major concern.

[22] It is also true that bank 1 would be more likely to achieve a large net sender position, and thus be a liquidity provider. But the free-riding measures are asymmetric since L_i^s and Λ_i^s cannot fall below zero, so the two effects do not cancel out.

We therefore regress the observed free-riding measures on these four independent variables:[23]

- proportion of values sent to all other banks in the system
- proportion of volumes received from all other banks in the system
- average size of payments sent
- average size of payments received.

We would expect values sent to have a negative coefficient, since banks which send more are heavier users of liquidity and so more likely to be identified as free-riders. Conversely, we would expect volumes received to have a positive coefficient, since for a fixed average payment size, higher volumes received would imply a lower probability of achieving very positive or very negative liquidity positions (by the law of large numbers).

We would expect average size sent to have a negative coefficient, because an increase in this variable while holding the others constant implies a greater likelihood of achieving more extreme positions on the random walk, as described above.

Finally, we would expect average size received to have a negative coefficient, because increasing it while holding total volumes constant implies that the bank receives a greater total value of payments over the course of the day, meaning it uses less of its own liquidity.

3.3.2.3 Results

Table 3.3 shows the results for the model where the y_j are the free-riding on cost measures (FRC) or free-riding on risk measures (FRR) observed over this period.[24]

---

[23] An alternative model might have as independent variables the values/volumes sent and received against *each* of the other banks in the system, rather than the total. But for a system with twelve banks this would mean $4 \times (12 - 1) = 44$ independent variables in the regression. This model would sacrifice parsimony for accuracy. Moreover, it would not be applicable to a system with a different number of banks.

[24] The Machado-Santos Silva (2000) test for heteroskedasticity rejects the null. We therefore estimate standard errors which are asymptotically robust to heteroskedasticity, using the qreg2 Stata command.

Table 3.3 **5% quantile regression of free-riding measures against values and volumes sent and received excluding Bank of England and CLS Bank, 1 January to 31 May 2010**

Dependent variable	FRC	FRR
Value sent	–1.17161***	–1.00826***
	(0.00328)	(0.00074)
Volume received	0.22027***	0.01015***
	(0.00367)	(0.00086)
Average size sent	0.00119***	0.00006***
	(0.00007)	(0.00001)
Average size received	0.00091***	0.00003**
	(0.00008)	(0.00001)
Constant	–0.00357***	–0.00016***
	(0.00005)	(0.00001)
Pseudo R^2	29.07%	7.62%
N	244,800	244,800

Standard errors are given in parentheses. Significance at 10%, 5% and 1% levels is denoted by *, ** and *** respectively.

In each of the two models, all four independent variables are non-zero. The coefficient for values sent is close to –1 in each case, which is unsurprising given the expressions for both free-riding measures explicitly include this term. This means that a bank which takes a 1% greater share of system-wide values sent will reduce its free-riding scores by a little over 0.01 (ie it will appear to be free-riding more).

Volume received has a positive coefficient, as expected. But the magnitude of the coefficient is small in the case of free-riding on risk r_i^s. Average values sent and received do not have the expected sign: in all cases the coefficients are positive. Although these are statistically significant, they are not economically large. The range of observed average values sent is 5.31, so variations in this regressor would make a difference of at most 0.0063 to the free-riding on cost score c_i^s and at most 0.0003 to the free-riding on risk score r_i^s. The range of observed average values received is 5.84, leading to similarly small effects.

Total liquidity actually used by these twelve banks averaged £18.7bn each day over this period, while in the simulations the system required £17.9bn on an average day. These results suggest that, in aggregate, the simulations do not differ significantly from reality. That is, strategic delay and urgent payments do not appear to result in a much less efficient system compared to the case where payments are simply made randomly.[25] However, at the level of individual settlement banks there may still be discrepancies between shares of liquidity usage and liquidity provision.

We can compare the free-riding measures to the actual liquidity used by each bank. The results exhibit a wide range, as shown in Table 3.4 below. We count the number of occasions when y_j exceeds $X_j'\hat{\beta}$ – in other words, the observed amount of free-riding exceeds that which the characteristics of the bank would suggest, at a 5% confidence level.

Table 3.4 **Comparison of actual free-riding measures over the period 1 January to 31 May 2010 to the 5% threshold values produced from the simulations**

	FR on cost	FR on risk
Threshold value exceeded, across all banks	18%	20%
Number of banks which never exceed the threshold	4	1
Number of banks which exceed the threshold on more than half of occasions	2	3

Across all banks, the threshold value is exceeded on around one-fifth of occasions for each measure, when 5% would be expected. This suggests that some banks do appear to free-ride, either for structural or strategic reasons.

3.4 Concluding remarks

Our cost and risk-based measures allow us to evaluate the liquidity that banks provide in RTGS systems compared to their share of liquidity usage, and examine whether this has changed over time. However, they cannot determine whether under-provision is strategic. The random walk analysis attempts to address this by simulating the payments system, abstracting from intraday payment timing decisions. We find that banks' liquidity usage varies with both value and volume, but that aggregate liquidity usage is not much more than

[25] Note that the simulations do not mimic an optimally efficient system. That would be a system where banks coordinate their payments so that large net sender positions are never built up.

might be expected in a system where payments are drawn from the original distribution in a random order. This suggests that strategic delay does not have a substantial impact on the system as a whole, though it may be more significant at the level of individual settlement banks.

We develop and calibrate a model which allows us to predict expected liquidity usage for a bank of a particular size. This could be useful for monitoring behaviour in the system and identifying banks which appear to be providing less liquidity to the system than they use. And it could be used to help in the monitoring and calibration of liquidity regulation. Suppose settlement banks are required to hold buffers of liquid assets in order that they can make payments in an orderly fashion even in a stress scenario. If these buffers are continuously calculated based on past activity, banks may have an incentive to delay their payments so that their regulatory buffer will be reduced at subsequent recalibrations. Our model could help to detect this behaviour, and also could be used to calibrate buffers independent of strategic actions.[26] Of course, any such calibration should take into consideration that intraday timing payment can vary for structural as well as behavioural reasons.

[26] For a more detailed discussion of this issue, see Ball et al (2011).

References

Ball, A – Denbee, E – Manning, M – Wetherilt, A (2011) **Intraday liquidity: risk and regulation.** Bank of England Financial Stability Paper, No. 11.

Bank of England (2011) **Financial Stability Report.** December.

Benos, E – Garratt, R – Zimmerman, P (2012) **Bank behaviour and risks in CHAPS following the collapse of Lehman Brothers.** Bank of England Working Paper, No. 451.

Committee on Payment and Settlement Systems (1993) **Payment systems in the Group of Ten countries.** Bank for International Settlements, December.

Galbiati, M – Giansante, S (2010) **Emergence of networks in large value payment systems (LVPSs).** DEPFID Working Paper No. 01/2010.

Jackson, J P – Manning, M J (2007) **Central bank intraday collateral policy and implications for tiering in RTGS payment systems.** The Future of Payment Systems, Routledge.

Koenker, R – Bassett, G Jr (1978) **Regression quantiles.** Econometrica, Vol. 46, No. 1, 33–50.

Leinonen, H – Soramäki, K (1999) **Optimizing liquidity usage and settlement speed in payment systems.** Bank of Finland Discussion Paper 16/1999, Helsinki.

Machado, J A F – Santos Silva, J M C (2000) **Glesjer's test revisited.** Journal of Econometrics, Vol. 97, 189–202.

Manning, M – Nier, E –Schanz, J eds. (2009) **The economics of large-value payments and settlement.** Oxford University Press.

Norman, B (2010) **Liquidity saving in real-time gross settlement systems – an overview.** Bank of England Financial Stability Paper, No. 7.

Soramäki, K – Bech, M L (2001) **Gridlock resolution in interbank payment systems.** Bank of Finland Working Paper No. 9.

Valukas, A R (2010) **Lehman Brothers Holdings Inc. Chapter 11 Proceedings examiner's report.** U.S. Bankruptcy Court Southern District of New York, Jenner and Block LLP, Section III.A.5.

Chapter 4

Is this bank ill?
The diagnosis of doctor TARGET2

Ronald Heijmans – Richard Heuver**

* Heijmans and Heuver can be reached ronald.heijmans@dnb.nl and richard.heuver@dnb.nl, respectively. We thank Ron Berndsen, Hans Brits, Iman van Lelyveld and the referees of the Bank of Finland for providing useful comments. The views expressed are those of the authors and do not necessarily represent the views of De Nederlandsche Bank.

4 Is this bank ill? The diagnosis of doctor TARGET2

Abstract

We develop indicators for signs of liquidity shortages and potential financial problems of banks by studying transaction data of the Dutch part of the European real time gross settlement system and collateral management data. The indicators give information on 1) overall liquidity position, 2) the interbank money market, 3) the timing of payment flows, 4) the collateral's amount and use and 5) bank run signs. This information can be used both for monitoring the TARGET2 payment system and for individual banks' supervision. By studying these data before, during and after stressful events in the crisis, banks' reaction patterns are identified. These patterns are translated into a set of behavioural rules, which can be used in payment systems' stress scenario analyses, such as eg simulations and network topology. In the literature behaviour and reaction patterns in simulations are either ignored or very static. To perform realistic payment system simulations it is crucial to understand how banks react to shocks.

4.1 Introduction

The financial crisis, which erupted in the United States in the summer of 2007, clearly showed the mutual dependence of the banking system on a worldwide scale. The crisis intensified after the failure of Lehman Brothers in September 2008. As a result of this failure the lending in the interbank market decreased significantly, see eg Heijmans et al (2010) for the Dutch market, Guggenheim et al (2010) for the Swiss market and Akram and Christophersen (2010) for the Norwegian market. As banks grew very reluctant to lend money to each other, central banks became worried that the interbank market would dry up completely. To prevent this from happening central banks worldwide, including the European Central Bank (ECB), responded with both conventional and unconventional monetary policy.[1] Studies on the structure of the interbank market can be found in Bech and Atalay (2008), who studied the topology of the federal funds market, Wetherilt et al (2010), who looked at the sterling unsecured loan market during 2006–2008, and Imakubo and Soejima (2010), who studied the microstructure of Japan's interbank money market.

A bankruptcy or a bail out usually does not come as a complete surprise. Before a bank goes bankrupt or has to be bailed out by the government or commercial parties in order to survive there are often rumours about its financial soundness. When time progresses these rumours might even become clear facts. The failure of Lehman obviously shocked the market, such a large bank to go bankrupt and not be saved by the government. Besides the failure of Lehman Brothers many other news facts shocked the market, such as failures of smaller banks, nationalisations of systemic important banks, state support etc. These news facts impacted the market perception of the troubled bank, which can become visible in the interbank money market (higher interest rates and lower borrowing volume) and in delays in payments by and to the troubled bank. The changed behaviour of one or more banks can consequently prompt many other banks to change their behaviour. This might in extreme cases lead to a total gridlock in which everyone is waiting for someone else to make the first payment. Such a situation will not only affect the payment system but can also jeopardise the financial system as a whole. If a bank intends to delay payments or change interbank interest rates based on rumours (not facts) it needs to trade off two kinds of risks: liquidity and reputation risk. A negative intraday position or outstanding loans vis-à-vis a counterparty is risky if one party is worried about its counterparty's ability to meet its obligations to that party. Delaying payments based on incorrect information can damage the debtor's reputation.

The research question of this paper is how to identify liquidity problems of a bank using Large Value Payment System (LVPS) transaction and collateral data. The literature focusses mainly on developments in the (unsecured) interbank money market, using an algorithm to identify interbank loans from the LVPS transaction data, see beginning of this section. This paper looks at all main liquidity influencing elements and actors behind these elements visible in the payment system. Besides, this paper transfers behavioural changes

[1] The unconventional monetary policy measures of the ECB consist of very long-term tenders (maturity up to 1 year) and the purchase of covered bonds.

found in the Large Value Payment System (LVPS), TARGET2-NL, and collateral data into a set of behavioural rules, which can be used in scenario analyses. To answer the main question we first look at the overall liquidity position of a bank. The overall development of the liquidity position provides an overview of a bank's liquidity streams and how it funds itself. After the overall liquidity position we look into more detail at the developments of the funding patterns. The developments in the interbank money market can eg show that one or more banks have difficulties obtaining liquidity out of this market to fund their liquidity needs. It is also possible that a bank or group of banks becomes more reluctant to lend their surplus liquidity to other banks as a result of increased market stress. If a bank can obtain insufficient liquidity from the interbank market it also has the option to borrow from the ECB (secured by its collateral) or use collateral to obtain intraday credit. If the use of collateral intensifies this can be a signal of near future liquidity shortages. If a bank faces difficulties funding itself on the interbank market and cannot obtain more liquidity from the ECB (based on the amount of its collateral) and as a result has no more liquidity available to make payments, it has no other option but to delay payments until it has received payments. A delay in outgoing payments can be a liquidity shortage signal. When the market suspects (serious) liquidity/financial problems with a certain counterparty it can delay some of their payments to this counterparty until it has received liquidity from this counterparty. A delay in incoming payments can be a signal that the market perceives this bank as more risky. If a bank's problems persist and its customers might loose faith in their bank at some point and transfer their funds to another bank or withdraw cash at the ATM, a bank run is born.

Figure 4.1 **Overview of liquidity management elements**

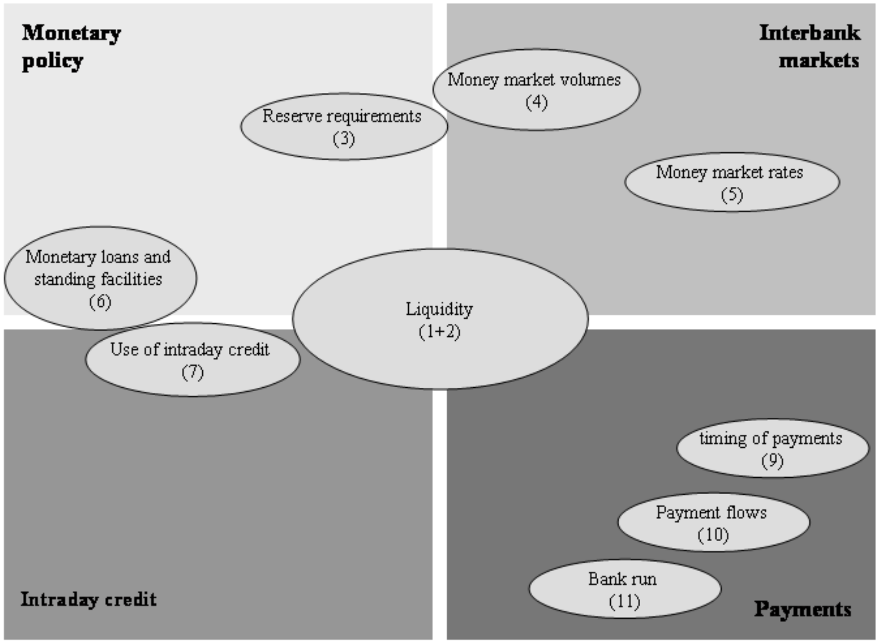


Figure 4.1 gives an overview of the areas of interest with respect to the liquidity management of a participant in the European Large Value Payment System (TARGET2). First there are the 'real' payments (bottom right-hand panel). These are the day-to-day payments on behalf of a customer or themselves which have no link with funding transactions or transactions to and from the ECB. A second area is the interbank money market (top right-hand panel), which gives information on interest rates and volume developments. Another area is the ECB facilities (top left-hand panel), which provide information on the use of tenders and how banks fulfil the ECB's reserve requirements. Finally, the bottom left-hand panel is the collateral amount and use. Collateral can be used for long-term loans (tenders), overnight credit (marginal lending) and intraday credit. The numbers in the overview correspond to the order the figures are presented in Section 4.3.

The changes in reaction patterns can be used to closely monitor the payment system and as a support tool for the supervision of banks to get up-to-date information on the current status of individual banks. As the payments data are available the next day, for most purposes the information on potential changes in behaviour is very timely.[2] A set of banks' behavioural rules is developed by studying the transaction data before and during the crisis. This set of rules can be used in e.g. scenario analyses of payment systems or in developments of the network structure after a disruption. This paper only looks at (potential) liquidity problems or 'illnesses' based on transaction and collateral data of participants in the LVPS and does not purport to say anything about the underlaying causes (eg a risky business model or bad management of a bank).

Relevant literature in the context of this paper is as follows. Banks are used to 'liquidity shocks' arising from unexpected changes in liquidity demand. Allen and Gale (2000) distinguish between two types of uncertainty. First, the idiosyncratic uncertainty, which arises from the fact that for any given level of aggregate demand for liquidity there is uncertainty about which banks will face that demand. The second type of risk concerns the aggregate uncertainty that is due to the fact that the overall level of the demand for liquidity that banks face is stochastic. These unexpected liquidity fluctuations impact the smooth operations of payments and Real Time Gross Settlement systems (RTGSs), besides affecting the banks' liquidity management (Iori et al, 2008), see for a description of LVPSs (and RTGSs) Section 4.2. Banks use the interbank money market (both secured and unsecured) to solve temporary shortages on their account. Cocco et al (2009) show that relationships are important for the banks' ability to access interbank market liquidity. The bilateral nature of this market enables banks to establish such relationships. Apart from access to liquidity, relationships do matter for both smaller and larger banks in negotiating favourable when borrowing and lending terms (Cocco et al, 2009, Carlin et al, 2007). We expect that relationship also plays a role in banks' payment behaviour and that banks are sooner inclient to delay payments if they expect a problem with a counterparty they do not have a relationship with. This can be inferred by the fact that banks do not want to be known, especially by one of their 'friends', as the one that pushed you over the edge of bankruptcy.

McAndrews and Rajan (2000), McAndrews and Potter (2002) and Bech and Garratt (2003) argue that the decisions made by banks in the U.S. LVPS Fedwire can be interpreted as a coordination game. Bech and Garratt (2006) have developed a stylised game theoretical model in which the timing of payments is reduced to two time periods: morning (in time) and afternoon (delayed). Abbink et al (2010) have

[2] Real-time monitoring is performed by the operators of a Real Time Gross Settlement.

conducted an experimental game based on their theoretical model. In this game they investigate how the behaviour in the payment system is affected by disruptions. Their main findings are that when the equilibrium of the payment system moves to the inefficient one (delaying payments) it is not likely that the behaviour moves back to the efficient equilibrium (paying in time). Besides, coordination on the efficient equilibrium turns out to be easier in a market with a clear market leader. Lastly, they find that small disruptions in coordination games can be absorbed easily, but when frictions become larger the system quickly moves to the undesired equilibrium and stays there. The fact that the payment system can be seen as a game illustrates that behaviour plays a role in these systems, especially in stressed times which change this game's dynamics.

Koponen and Soramäki (2005), Bech and Soramäki (2005), Ledrut (2007) and Heijmans (2009) are examples of payment systems' simulations, based on historical data. The behaviour of banks in these papers does not represent realistic behavioural patterns. A description in networks' terms, see eg Soramäki et al (2007) or Pröpper et al (2008), gives information on the critical participants and the level of dependencies between the participants in the payment system. It does not give information on how participants behave. Before the crisis started there was not much empirical evidence on how banks behave in times of stress, as there were not many stressful events. The 9/11 attacks gave some insight but there were no banks at that time facing severe liquidity problems over a longer time or that went (almost) bankrupt, see Lacker (2004). In order to improve the realism of simulations and the dynamics of network structures it is essential to include behaviour into the analysis.

The outline of the paper is straightforward. Section 4.2 describes the data set used for the analysis. Section 4.3 describes how the TARGET2 transaction and collateral management data can be used to find signs of liquidity shortages of individual banks. Section 4.4 describes the set of behavioural rules based on evidence found during the crisis, and Section 4.5 concludes and gives policy recommendations.

4.2 Large value payment systems

Before we move on to the identification process of liquidity problems of banks and the behavioural rules find in the LVPS data and collateral management data we first give a description of this LVPS.

Large value payment systems (LVPS) play an important role in the economy. With the help of these systems, banks can settle their (customer) obligations immediately and irrevocably. The irrevocability of the payment is very important in the LVPS as the receiving bank can reuse the liquidity without running the risk that the liquidity has to be repaid to the sending bank in case of bankruptcy of that bank. Because of their economic relevance, LVPS have to live up to high standards. They must comply with the core principles which were developed by the central banks, co-ordinated by the Bank for International Settlement (CPSS, 2001). The most important euro-LVPS is TARGET2.[3] Another system that is used is EURO1.[4]

4.2.1 TARGET2

TARGET2 is the large value payment system of the eurosystem, which is used to execute time-critical payments. Besides the euro countries, there are 6 non-euro European countries that are connected to TARGET2 for the settlement of euro payments.[5] Technically, it is a centralised system, which means that there is one platform for all participants to settle their payments. Legally, TARGET2 is a decentralised system. Each country still has its own legal documentation. The conditions are, however, maximally harmonised, but small deviations are allowed if required by national legislation. In a legal and business sense, one of the central banks is the intermediary channel between a financial institution and TARGET2.

TARGET2 can only be used by institutions which meet the access criteria. The most important institution types that can gain access to TARGET2 are credit institutions established in the European Economic Area (EEA), EU member states' central banks including the ECB and central or regional governments treasury departments of member states active in the money market. Most other financial firms, non-financial firms and consumers do no meet the access criteria of TARGET2

All payments executed in TARGET2 are stored in a datawarehouse. The average daily turnover in TARGET2 in 2010 was EUR 2,267 billion, which corresponds with an average number of transactions of more than 340,000. The Dutch part of TARGET2,

[3] TARGET2: Trans-European Real Time Gross settlement Express Transfer.
[4] EURO1 is a private sector owned payment system for domestic and cross-border single payments in euro between banks operating in the European Union.
[5] Bulgaria, Denmark, Estonia, Latvia, Lithuania, Poland and Romania (status July 2011).

TARGET2-NL, accounts for 13% and 10% respectively. For a more detailed description of large value payment systems see Heijmans et al (2010, Section 2).

The Dutch market is characterised by a few large banks and many small(er) ones. In TARGET2-NL there is also a few large British banks.

4.2.2 Description of the data

Financial institutions settle various types of payments in TARGET2, such as payments on behalf of a customer, bank-bank payments, payment of the cash leg of a security transaction, pay-in of CLS (continuous linked settlement) to settle foreign exchange transactions, and so on. The data used in our analysis contain transaction level data of TARGET2-NL (and its predecessor the Dutch RTGS TOP) between 1 January 2005 and 28 February 2011 and the collateral's amount and use of each individual bank in TARGET2-NL (and TOP).[6] The accounts of De Nederlandsche Bank and the Dutch Treasury (including its agency) were excluded, as these institutions are no commercial banks.

4.3 Monitoring individual banks

Besides the Lehman Brothers' failure several stressful events occurred in the Dutch financial system, like the nationalisation of Fortis – ABN AMRO, the bankruptcy of DSB Bank and the state support of several larger financial institutions. The effects of these events have to a lesser or greater extent become visible in the TARGET2 transaction and collateral data. These data have been investigated to develop a monitoring tool for both individual banks and the market as a whole. This section describes the monitoring tool. For illustration purposes the information in the figures is based on TARGET2-NL data. Most graphs presented in this section could be shown for a single bank, group of banks or the whole market. Section 4.4 translates the visible effects into a set of behavioural rules.

The monitoring tool consists of several different indicators. If only one indicator changes this need not signal a problem, but if there are

[6] TARGET2-NL was launched on 18 February 2008.

more indicators heading in a certain direction this may be a sign that there are serious liquidity and/or financial problems with a bank. We have seen in the data several banks with only one indicator showing potential liquidity problems while the other indicators were neutral. However, with the banks in trouble (like Fortis and DSB) there were several indicators heading in the direction of liquidity problems. This is comparable with the medical doctor's differential diagnoses. The doctor diagnoses the patient's problem by asking the patient questions. In some cases there might not be a medical problem at all or only a light flu which will be over in a few days without treatment. But if more indicators point to a serious disease, immediate treatment might be necessary. The differential diagnosis of the central banker is based on the changing reaction patterns obtained from e.g. the TARGET2 and collateral data. By combining the different elements we can see whether a bank (patient) has or may have serious liquidity or financial problems (illness) and whether action from the central bank or supervisor authority is required.

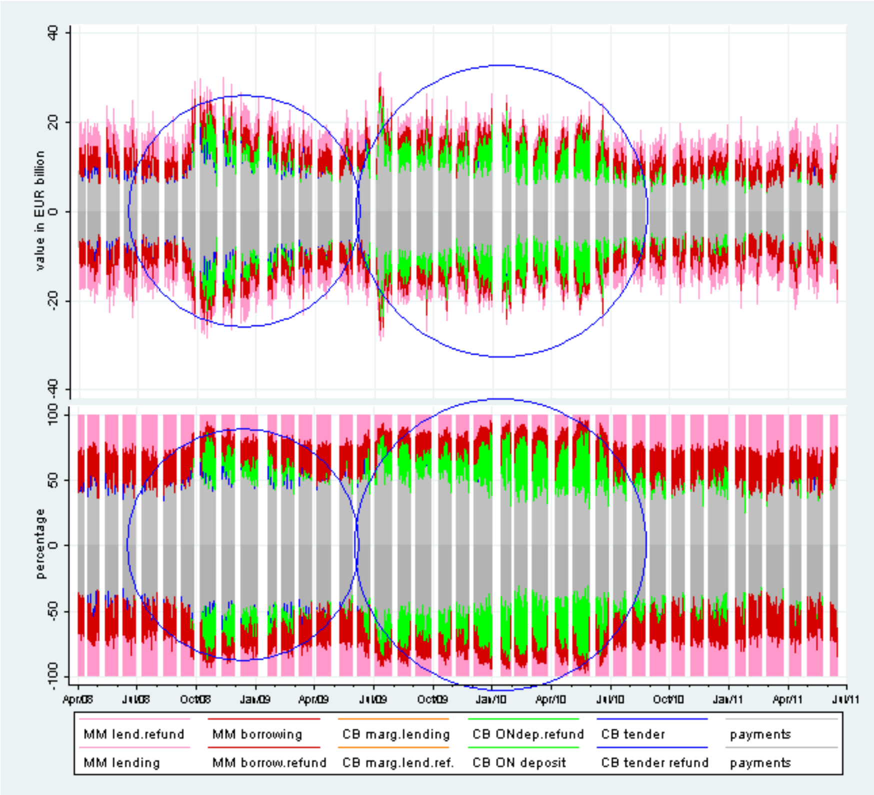
Some caution is in order for banks which are not very active in the payment system or in the interbank market. For Lack of sufficient payment transactions the monitoring tool provided in this section might give misleading information in such cases.

4.3.1 ECG of liquidity

When a medical doctor examines a patient, he often listens to the heart. If the doctor hears abnormalities he may decide to make an electrocardiograph (abbreviated to ECG). The ECG is the transthoracic interpretation of the heart's electrical activity over time captured and externally recorded by skin electrodes. From the ECG a lot of information can be obtained on the physical state of a patient. Likewise we seek information on the 'health' of banks. Supervisors want to have information on both the solvency (long-term) and liquidity (short-term) position.[7] Also operators of LVPSs want to know if a bank faces liquidity or technical problems, as such problems may affect other banks in the payment system. An additional complicating factor for supervisors is that the ECG of liquidity, in contrast to patients, can be very different for each bank depending on their liquidity management and business characteristics.

[7] It is not possible to say anything about the solvency of a bank based on TARGET2 transaction or collateral data.

Figure 4.2 **ECG of liquidity: flow of payments (top panel), relative flow of payments (bottom panel), 18 February – 30 April 2011.**



The top 6 elements of the legend belong to the positive and the bottom 6 to the negative vertical axis.

4.3.1.1 ECG of liquidity: Payment flows

Figure 4.2 shows the most important payment types' liquidity flows of a bank (1st panel shows absolute flow values and 2nd panel the relative values), distinguishing payments on behalf of a customer or themselves or 'real' payments, monetary policy, standing facilities (deposit, marginal lending) and money market lending and borrowing.

The figure shows several maintenance periods.[8] The white vertical gaps in the figures (and also in the graphs in the rest of this chapter) represent the separation between two maintenance periods. The positive values refer to the incoming values and the negative values refer to the outgoing ones. The reason for choosing the maintenance period is that there are strong cyclical patterns in eg money market transactions which are used to level surplus or shortages, see Heijmans et al (2010). They have found eg that at the end of the maintenance period there is a significant increase in money market transactions. The secured lending is not separately visible in this graph as it is not possible to identify them (reliably) from the TARGET2 transaction data.

Figure 4.2 contains the following information. This figure first gives an overview of the value of the most important payment flows in both the absolute (top panel) and the relative (bottom panel) sense. Especially the activity at the interbank market is interesting from a risk perspective (see also Section 4.3.2.2). If a participant's loses trust in the market and as a result is not willing to lend its (full) surplus, this will become visible by an increased use of the ECB's overnight deposit. On the other hand if a participant is unable to borrow the required liquidity at the interbank market, use of both tenders and marginal lending are likely to increase, see Section 4.3.2.3 for a detailed view of the use of ECB facilities. In other words the activity at the interbank market gives information on how banks perceive the risk of the market and of individual participants. The figure also shows the development and volatility over time of the several payments flows. Moreover, the figure gives information on the potential demand for liquidity. A bank with a lot of real payments (grey bars) has potentially more need for liquidity (tenders, interbank loans) in the absolute sense. A bank with a higher day-to-day volatility in the real payment flows potentially needs relatively more liquidity (tenders, interbank loans) while also its future liquidity position is more difficult to predict. As a modified example, the circles in Figure 4.2 show two changes which can be identified from this graph.[9] In the left circle there is a shift from interbank lending to the overnight deposit's use and the right circle shows a period the bank uses the overnight deposit extensively. At the right part of this circle there is a

---

[8] A maintenance period is the time frame in which at the end of the business day banks must maintain an average level of funds specified by the central bank. If a bank does not meet the maintenance requirement it will receive a penalty.

[9] The data has been modified for confidentiality reasons. These data do however reflect what has been observed in the data.

decrease again in overnight deposit and an increase of interbank lending and borrowing. This may signal an increased trust in the market circumstances.

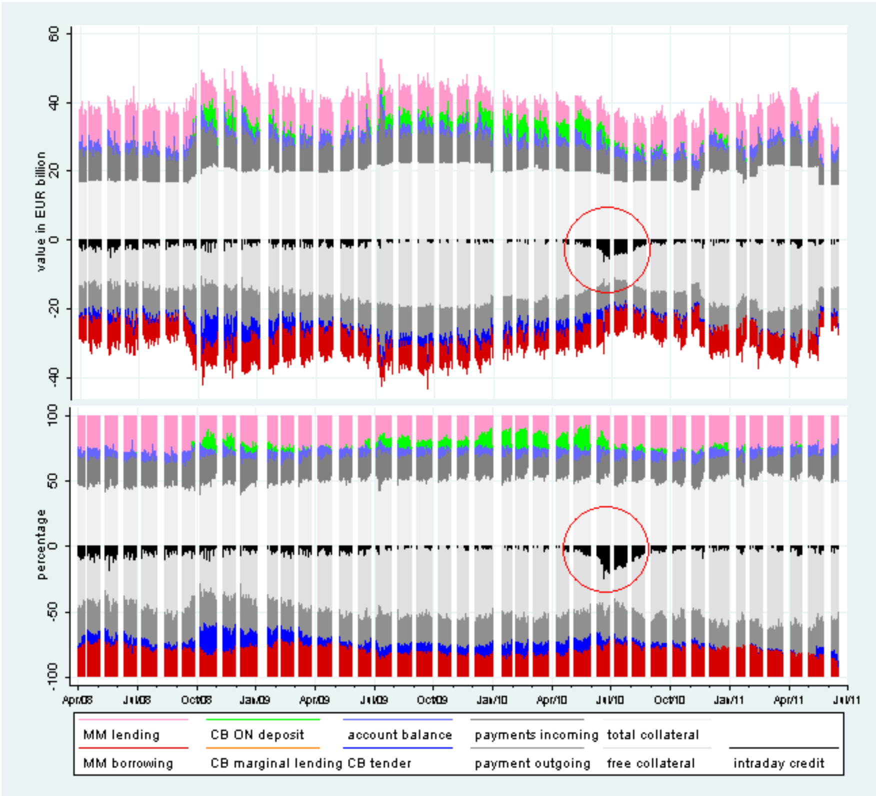
4.3.1.2 ECG of liquidity: outstanding values

Figure 4.3 shows the outstanding value of the payments flows. The positive vertical axis represents the assets of a bank, including its pledged collateral at the central bank, its total daily incoming payments, its outstanding interbank lending, its overnight deposit, its account balance and its incoming payments. The negative vertical axis on the other hand represents the liabilities of a bank, including its interbank borrowing, marginal lending, tenders, total daily outgoing payment, free available collateral (which can be seen as equity) and use of collateral for intraday credit. In other words Figure 4.3 presents a balance sheet of assets (positive vertical axis) and liabilities (negative vertical axis).

To illustrate the difference between Figure 4.3 and 4.2 we take the year tenders that the banks were able to obtain from the ECB in July 2009. The year tenders are visible in Figure 4.2 as two individual transactions one year apart: one incoming in July 2009 and the other outgoing July 2010. In Figure 4.3 the year tender stays visible the whole year starting in July 2009 until July 2010. At the same time the free collateral will decrease/increase with the exact amount of the tender at the start/end of this tender (assuming a bank does not change the amount of collateral at the start or end of the tender). The same reasoning is true for interbank loans (lending bars and borrowing bars), except that there is no effect on the free collateral.

Figure 4.3 contains the following information. First, information about a bank's differences in the funding sources from interbank money market to the ECB or vice versa. If a bank moves its funding from the interbank market to the ECB, this may indicate that a bank has difficulties in funding itself in the market. Furthermore, the figure makes clear whether a bank is a lender or a borrower and how lending and borrowing changes over time. The figure also shows changes in the amount and use of collateral. If the amount of collateral e.g. decreases, the ability of a bank to withstand (new) shocks decreases along. This is especially true for banks which have a relatively low amount of collateral relative to their payments. Besides the shifts from lending/borrowing to overnight deposit and vice versa which have also been identified by Figure 4.2, the circle in Figure 4.3 also shows as an example the period the bank suddenly used more intraday credit, which may be a sign of (near future) liquidity problems.

Figure 4.3 **ECG of liquidity: absolute outstanding value (top panel), relative outstanding value (bottom panel), 18 February – 30 April 2011.**



The top 5 elements of the legend belong to the positive and the bottom 6 to the negative vertical axis.

4.3.2 Demand and supply of liquidity

4.3.2.1 Reserve requirements

The ECB requires banks to hold a minimum cash reserve on average during the maintenance period. The main reason for banks to borrow liquidity is that they have to meet the ECB's requirements. Due to natural fluctuations banks face shortages and surpluses on a regular basis, see Allen and Gale (2000). If a bank expects that it will not meet its requirement it will use the interbank market or the ECB facilities to meet them. In order to see if certain changes in the ability to lend liquidity are worrisome it is necessary to know if a bank is in need of liquidity to meet its requirements. Figure 4.4 is an illustration of how a bank could or could not meet its cash reserve requirement. The black bars denote when the bank has a surplus and the grey bars when it has a shortage relative to its cash reserve requirement.

Figure 4.4 **Five stylistic examples of how a bank meets (I to IV) and fails to meet (V) its maintenance periods.**

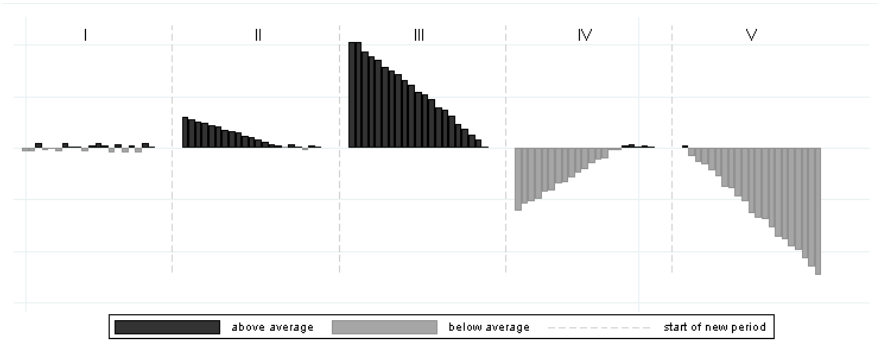


Figure 4.4 enables us to answer three questions with respect to cash reserve requirements: 1) Does a bank meet its reserve requirement? 2) When does a bank start to meet its reserve requirements? 3) Has the timing of meeting the reserve requirement changed over time? In maintenance period I of the figure, the bank steers the maintenance requirements on a daily basis. This means that the bank neither has a large surplus nor shortage on any business day during the maintenance period. In maintenance period II the bank starts with a surplus, which vanishes over the course of the period and ends in the last week with values close to the reserve requirements, like in period I. In maintenance period III the bank starts with a large surplus, which vanishes over the course of the whole period. In period IV the bank starts with a relatively large shortage. This shortage decreases as the maintenance period progresses. If a bank is unable to meet its requirements it can expect a penalty from the supervisors. In period V the bank's shortage relative to its maintenance requirements only increases. This will be the case if the bank is no longer able to solve its problems.

Figure 4.4 also helps to determine if a bank needs liquidity from the interbank money market. If a bank has a surplus relative to its reserve requirement (black at a certain point during the maintenance period), it does not need liquidity from the market but is able to lend liquidity to the market. Whether it will do the latter depends on its expected liquidity need for the rest of the maintenance period and on the perceived risk of counterparties which want to borrow from this bank.

4.3.2.2 Interbank lending and borrowing

If the market suspects liquidity or financial problems with a certain counterparty and has serious doubts whether that counterparty is able to fulfil its payment obligations, banks become more hesitant to provide liquidity. This will show up in the amounts banks are willing to lend to this counterparty and/or in the interest rate this counterparty has to pay. Figure 4.5 shows the amount of interbank overnight lending and borrowing over time by a bank. The algorithm used to filter the interbank loans has been developed by Heijmans et al (2010), which is based on the algorithm by Furfine (1999).[10] The developments of the lending and borrowing amount reflect the bank's market perception (lending) or the market's perception of this bank (borrowing).[11] A bank's lending and borrowing behaviour also gives information on the bank's type. Is a bank on average a lender, a borrower or a money broker (both lending and borrowing at the same day)? In case a bank is generally a lender it is better able to withstand liquidity shocks than a borrower, because it can decrease its lending amount. A borrower on the other hand becomes even more dependent on other banks' liquidity in case of liquidity shortages. The circle in Figure 4.5 shows an example of the developments of both the interbank lending and the interbank borrowing volumes, which may

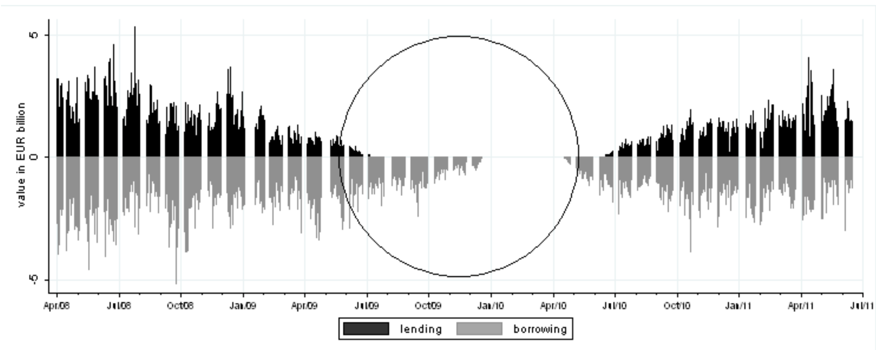
[10] Even though the algorithm can detect loans up to 1 year, only overnight loans are used. The reason for this is that most of the loans (in value) are overnight and no shift has been identified during the crisis from long-term to short(er)-term lending. Therefore the overnight loans are an reliable indicator of changes in the ability to lend at all maturities. The loans identified by Heijmans et al (2010) will mainly be unsecured. However, it is possible that the algorithm also detects (some) secured loans. This might happen if the liquidity is settled in TARGET2 and the securities are transferred in ESES free of payment (fop: security shift from bank A to bank B without having the payments on the security platform).

[11] Lending and borrowing are most likely predominantly unsecured. However, exact numbers on how much of the loans is secured are not known due to lack of information on securities cleared in other systems connected to a (secured) loan in TARGET2.

be seen as worrisome. First the lending amount decreases and later on the borrowing amount. If a bank has liquidity shortages it will stop lending. When the market suspects liquidity problems with a counterparty it will decrease or cease lending to this counterparty. To see whether the decrease in lending and borrowing is indeed a problem the information has to be combined with the use of ECB facilities and the way the bank meets its maintenance requirements.

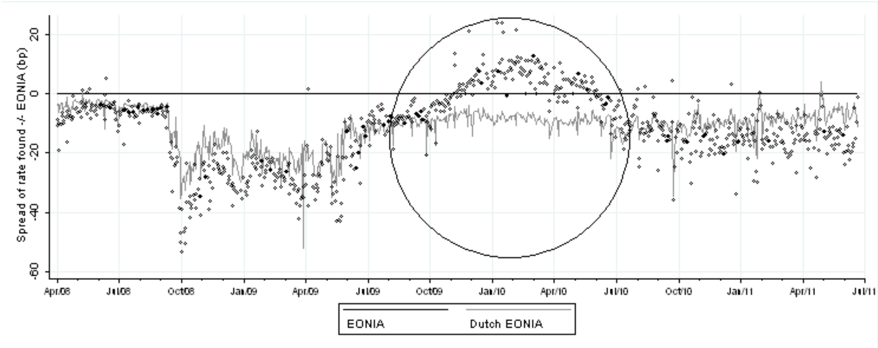
Figure 4.6 shows the development of bank borrowing rates relative to the Dutch average (or: the 'local' average) and the European average (EONIA).[12] If the market perception towards a single (or group of) participant(s) changes for better or worse, the interest rate will decrease or increase as a consequence. The circle in Figure 4.6 illustrates how interest rate developments may signal potential bank's liquidity problems. First the interest rates increase as a result of an increased market's risk perception to this bank. In other words the bank has to pay a higher price for its loans. When time progresses these increased rates move slowly back to normal. If the interest rates for a single bank increase significantly as shown in this graph this may be a signal of near future liquidity problems as it becomes more difficult for this bank to fund itself. The data showed that the first signs of liquidity problems will appear in the interbank lending and borrowing rates and volumes.

Figure 4.5 **Developments of the interbank lending and borrowing volume, 18 February 2008 – 30 April 2011**



[12] Local average based on algorithm and EONIA based on quotes.

Figure 4.6 **The relative development of the interest rate of a single participant compared to Dutch average and EONIA (black zero line), 18 February – 30 April 2011.**



4.3.2.3 ECB facilities

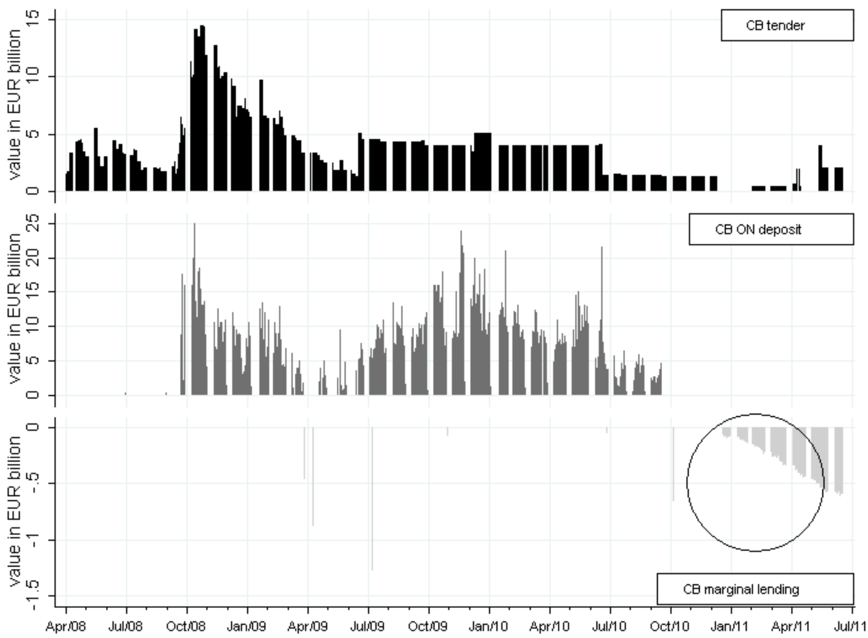
Besides the interbank market, banks can make use of the ECB tenders and standing facilities, see Figure 4.7. The tenders vary in duration from 1 week up to 1 year. A shift from eg interbank borrowing to ECB tenders can be a signal that a bank faces difficulties in fulfilling its obligations in the interbank market.[13] A bank's extensive use of the overnight deposit can reflect a lack of trust of this bank in its counterparties. However, during 2009 and 2010 the use of the overnight deposit was also (partly) caused by the excess liquidity of the ECB.[14] If a bank starts using the marginal lending intensively in a short time frame it is usually a signal of its inability to borrow from other banks. Especially the combination of a strong sudden decrease in the borrowed amount and/or a strong increase in interest rates and at the same time intensive use of marginal lending clearly signifies that a bank is having liquidity problems. This can be combined with a decrease in the amount of collateral, see Section 4.3.3.

To illustrate potential liquidity problems with the use of ECB facilities, the circle in Figure 4.7 shows an increase in both the amount

[13] During this crisis, the ECB year tenders where used intensively by many banks as a security measure to withstand potential future shocks and not necessarily because they were highly required in the short term.
[14] It is easy to check if the central bank has put too many tenders into the market when banks use both tenders and overnight deposits close to the tender's amount.

and frequency of marginal lending. Under normal circumstances, a bank would borrow on the interbank market, as this is the cheapest option. For the bank concerned, this was either not possible or did not yield sufficient funds to solve the liquidity shortage. We observed that only banks facing extreme liquidity shortages (just before being nationalised or collapsing) make intensive use of the marginal lending facility. Our data also showed that the overnight deposit facility was most often used by banks with a surplus not willing to lend this surplus.

Figure 4.7 **Use of ECB facilities by a bank (February 2008 – April 2011). The scale of the vertical axes of the 3 graphs are not the same**



4.3.3 Collateral

The central bank provides credit to banks by monetary policy (tenders), overnight credit (marginal lending) and intraday credit if banks meet the requirements for making use of the different types of credit. In order for a bank to obtain credit it has to be collateralised. The more collateral is available, the better a bank is potentially able to withstand temporary shocks. If, for example, a bank does not receive any or insufficient incoming payments, the balance at its account will decrease and eventually tend to drop below zero. Instead of delaying payments until it has received more payments, it can already fulfil its obligations by making use of intraday credit, secured by its collateral.

To monitor the use of the intraday credit for the whole market Figure 4.8 can be used. This graph shows the shares in the RTGS (this can be corrected for the weight a bank has in the total average turnover) of banks that have used a certain percentage of their collateral for intraday credit. The lightest grey part of the figure for example shows the banks' share (scale on y-axis is the interval [0,1]) which has used up to 30% of their collateral and the black part shows the share that has used up 90% to 100%. If, however, the figure shows less light grey and more black over time, this denotes that more banks use (almost) all their collateral for intraday credit and the payment system becomes more vulnerable to shocks given the amount of collateral. The circle in Figure 4.8 shows an example of a clear increase in the market's use of collateral for intraday credit.

Figure 4.9 shows the total available amount of collateral for one bank and the use for monetary loans, intraday credit and unused collateral for each day. The intraday credit value is the maximum amount during a business day. A few aspects of the total amount and use of collateral could be monitored: 1) the total available collateral's amount, 2) the collateral's use for monetary loans and intraday credit and 3) the change in the amount and use over time. The collateral's use for intraday credit is connected to the payments' timing when a bank has used (almost) all its free collateral for intraday credit. It will then soon be unable to make any payments before it has received incoming payments first. Especially if this bank expects or knows that it must make some very urgent and time critical payments it has to make sure that it has sufficient liquidity at its disposal for this purpose. It can do so by delaying less urgent and less time critical payments; see for the payment's timing Section 4.3.4. If a bank uses eg more than 90% of its collateral and it simultaneously shows signs of payment delays, this is a sign of (temporary) liquidity or financial problems. The circle in Figure 4.9 shows, as an example, a period of increased intraday credit use. In our data we have observed such an increase with banks facing difficulties funding themselves in the money market. Especially if these bank simultaneously face an increased outflow of liquidity.

Figure 4.8 **'Bush fire': Distribution of TARGET2-NL participants' collateral use for intraday credit**

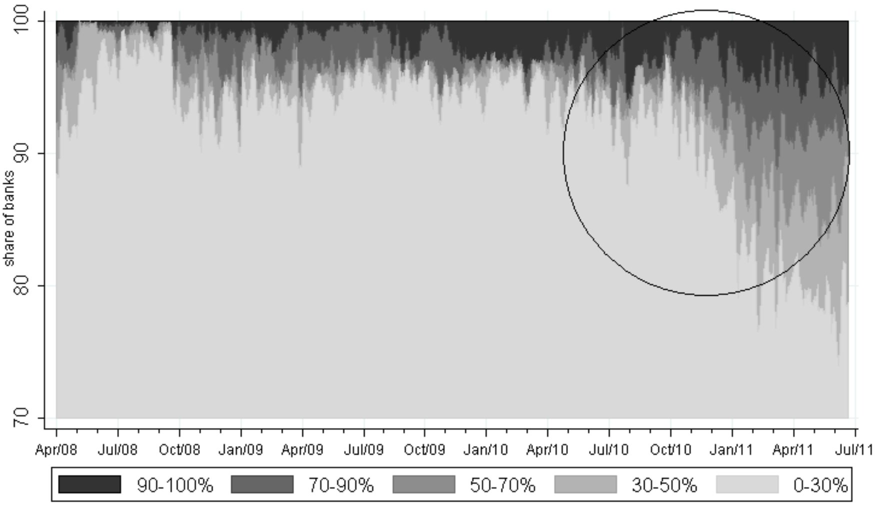
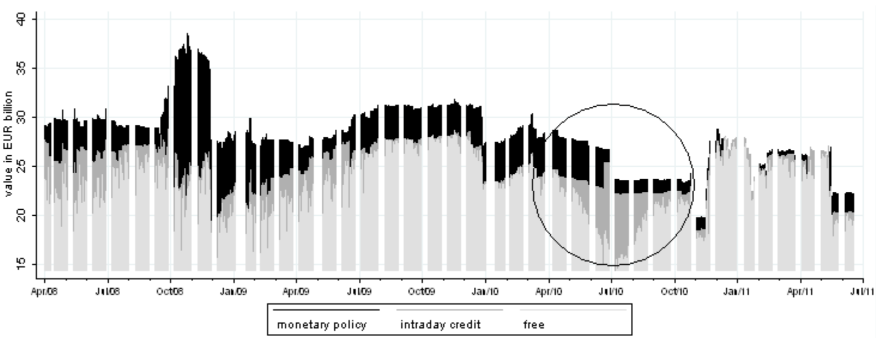


Figure 4.9 **Total amount and use of collateral pledged by commercial banks**



4.3.4 Timing of payments

4.3.4.1 Development of daily average timing of payments

Banks rely on each other's liquidity to be able to make their own payments. Due to the gross nature of RTGS systems the total amount settled on an average day are much higher than the available liquidity on the accounts of the RTGS system. If one or more (large banks) delay their payments this may affect other banks' liquidity position to the extent that they are also no longer able to make payments. The result would be a so-called gridlock, a situation in which everyone is waiting for someone else to make the first payment. The reasons why a bank delays its payments are 1) that it is technically not able to make them, as happened to some banks during the 9/11 attacks (Soramäki et al, 2007), 2) that it has no liquidity available to execute any payments or 3) that it delays intentionally. The reason why a bank delays is not important for the effects of this delay, but it does have an effect on the possibility to solve it quickly. This paper only focuses on either intentional delays or the inability to make payments as a result of liquidity shortages.

Figure 4.10 shows the value-weighted average incoming (black line) and outgoing (grey line) payments' time. The dark grey and light grey shaded areas are the corresponding 90% intervals. The graph distinguishes between bank-bank transactions excluding the interbank loan transactions and payments on behalf of a client (client transactions). The reason for excluding these loan transactions is that they differ in nature from 'real' payments. Besides, the current crisis has shown that the lending activity decreased significantly, see eg Heijmans et al (2010), without banks having any liquidity problems. As interbank loans are generally of high value their partial disappearance could change all payments' average timing significantly even though not a single payment was delayed or paid earlier. The client payments are usually not delayed as clients are assumed to have sufficient liquidity on their accounts.[15]

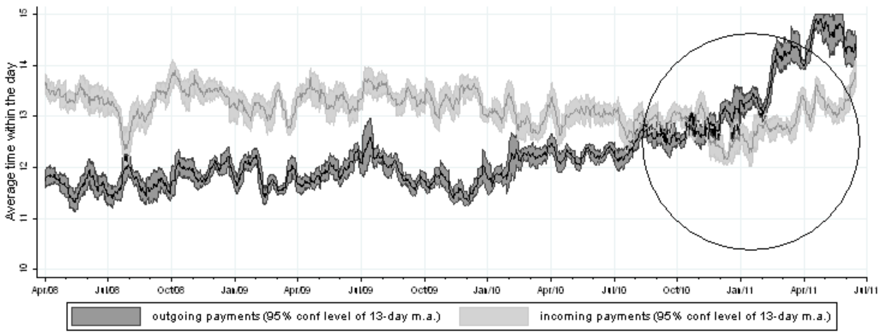
By way of an example, the circle in Figure 4.10 shows an increase in the average outgoing payments' time to a level above the average incoming payments' time, suggesting liquidity problems. For the weeks before Fortis was nationalised, our data show evident changes in the average payments' time. First there was a small increase in the outgoing payments' time, in response to which the market on average soon started delaying payments to Fortis. Fortis responded to this increase by reducing its average outgoing payments' time substantially. In other words: it began to transfer payments at a very early stage. Fortis continued to do so as long as permitted by its liquidity position.

An interesting feature of the graph is the change-over from an early payer (grey line below the black one) to a late payer, especially

[15] Anecdotal evidence obtained from liquidity managers of Dutch banks supports this idea.

for banks known to be the payment system's liquidity provider. Over the years, most large banks in the Dutch payments system are known to be early payers and provide the whole payment system with liquidity. The smaller banks use this liquidity to execute their payment obligations. If these large banks start to pay later or even become late payers this can seriously affect many banks' liquidity positions and in the worst case lead to a total gridlock. Large banks most likely provide so much liquidity to the payment system before receiving liquidity because: 1) their relative large cash reserve requirements enables them to make lots of payments before they have to use intraday credit and 2) as long as they receive their incoming payments directly effecting their payments, there is only a short time interval of credit risk (in case of a small bank's failure).

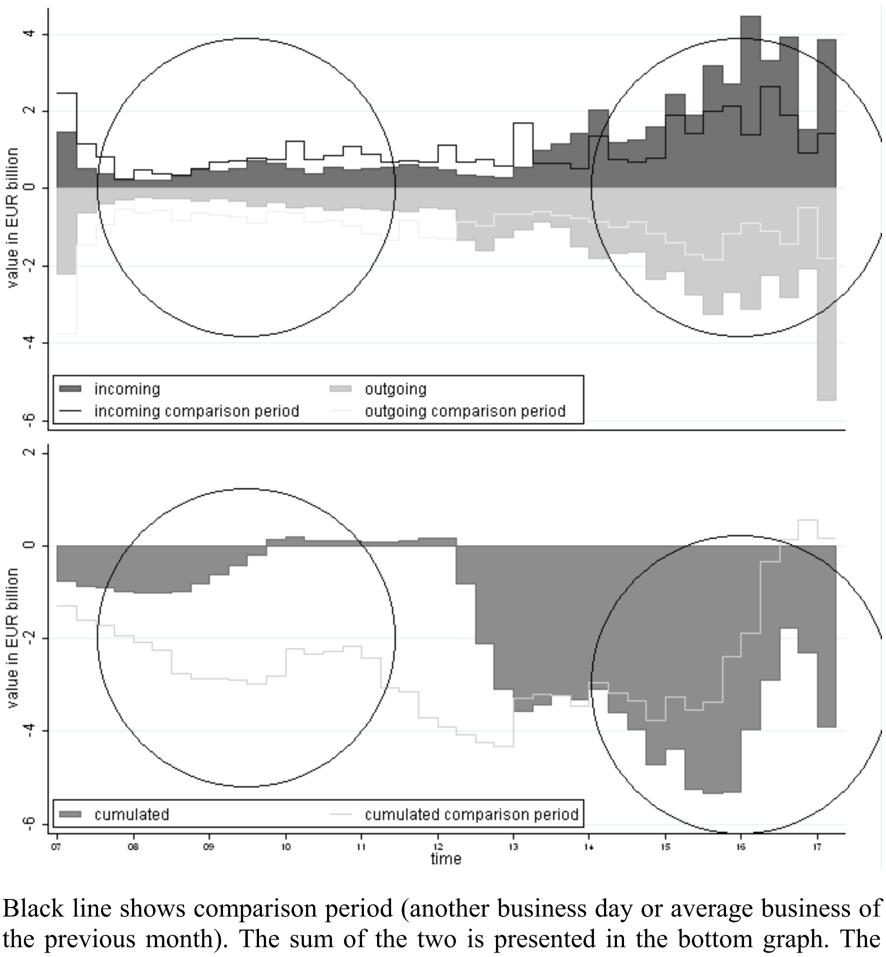
Figure 4.10 **Average time of incoming and outgoing payment flows (for February 2008 to February 2011)**



Top graph displays bank-bank transactions excluding interbank loans. The bottom graph shows the client transaction. The lines are the averages and the areas the 90% interval.

Within the scope of payment system monitoring it is important to know when a change of the average incoming or outgoing payments' time can be taken as a signal. The day-to-day fluctuation of Figure 4.10 is large and therefore not very meaningful for monitoring (noise). Looking at trend variations (difference with the previous days), during 7 or 14 days, will be more useful. When defining an automatic alarm signal a trade-off has to be made between the number of 'alarms' and the fact that there is something going on. The more often a change of timing causes an alarm, the more often the alarm will be false. However, if the change of timing chosen is very large there will hardly be any hits, which might result in missing important changes. An alarm in the change of timing can be shown by way of a traffic light. If the average incoming or outgoing time increases beyond eg the 95% interval once a yellow light can be given, and does so several times in a short period, a red light can be given. Anyhow, no matter what test or methodology is used there is always the probability of a false positive or false negative, similar to tests in the medical world. The outcome of a medical test can be positive, meaning you have the illness, but in fact you are not ill at all (false positive) or the outcome is negative but you are ill (false negative).

Figure 4.11 **Incoming and outgoing liquidity flows for each 15 minutes during a business day or average business day of a month (top graph).**



Black line shows comparison period (another business day or average business of the previous month). The sum of the two is presented in the bottom graph. The black line shows the comparison period.

If a yellow or red light is given for the average daily incoming or outgoing timing it is useful to zoom in on the intraday payment patterns to see if there have been specific times there were delays. Figure 4.11 shows the sum of the incoming and outgoing payment flows for every 15 minutes of the business day. The black lines show the comparison period to see the shift in timing. The bottom graph shows the cumulative balance up until that time of the day. In contrast to the average payments' timing (Figure 4.10), it is possible to see intraday developments and a comparison with a relevant other period. Figure 4.11 shows an example of a bank which has moved its outgoing payments to the end of the day. The incoming payments of the other have also been moved towards the end of the day.

4.3.5 Signals of a bank run

A bank facing liquidity and/or financial problems, which it is unable to solve, cannot keep these problems silent for the general public (businesses and consumers) forever. If this public starts losing faith in their bank and withdraw their money from this bank, the problems will accelerate and a bank run is born. This loss of faith may e.g. be triggered by a steep decrease in the stock price of a bank, as we have seen with Fortis, or a call for a bank run by an influential person as we have seen a few weeks before the DSB bankruptcy.[16]

A start of a bank run becomes visible in the data of TARGET2-NL in five ways: 1) the banknote withdrawals at the central bank, 2) the settlement of Equens (a Dutch settlement system), 3) client payments and 4) urgent customer payments (TNS). The first two bank run types can be seen as 'retail' bank runs since the net result of many customer transactions is settled in TARGET2. The last two types can be seen as the 'wholesale' bank runs since each individual transaction become visible in the TARGET2 data. An interesting aspect of each of the five bank run types is that the value settled is generally very low compared to the total daily payment by a bank (below a few percentage points). Nevertheless, by identifying the correct transactions of the bank run type from the TARGET2-NL transactions data a clear signal of a bank run can become visible as soon as this bank run starts.

[16] DSB used to be a non-systemic bank in the Netherlands.

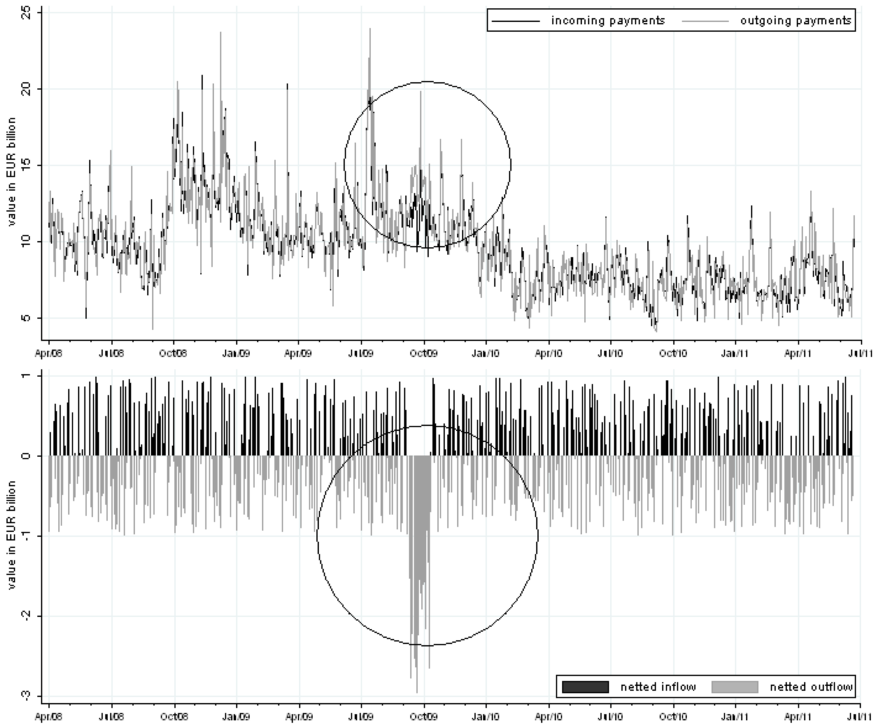
4.3.5.1 'Retail' bank run

The most well-known bank run is the cash bank run. Customers who have lost faith in their bank, stand in line of the their bank's ATM or counter to withdraw their savings. As demand for cash suddenly increases significantly, so does the commercial bank's demand for cash (bank notes). The bank can obtain the extra bank notes from their central bank, which brings bank notes into circulation.

Commercial banks' cash withdrawals and deposits at the central bank are visible in TARGET2 as individual transactions. The deposit and withdrawal patterns and the netted value differs from bank to bank, depending on the banks' clients which being more consumers withdrawing banknotes or retailers depositing banknotes. Besides the overall positive or negative flow there are clear seasonal effects, around eg Easter, Whitsun and Christmas. The reasoning behind these effects is that (foreign) consumers will hoard cash just before and during public holidays. The cash spent at retailers or the surplus finds its way back to a bank, which deposits the surplus at the central bank. If the public gets wind of a bank's possible failure it will suddenly and massively start taking out cash from the ATM or at the bank's counter, which is the start of a bank run. A recent example of a bank run was seen at Northern Rock in September 2007, see eg Shin (2009). Savers of this bank formed long queues to withdraw their life savings.

A second retail bank run can be told from the data related to settlements in TARGET2-NL by the settlement organisation (in the Netherlands this is Equens). Equens settles the debit card transactions in the Netherlands in multiple cycles per day. The net position of all debit card transactions is calculated and settled in TARGET2-NL. If customers of a bank suddenly spend much more of their money this will become visible as a strong negative position for this bank. Just as for cash deposits and withdrawals, whether a bank has a positive or negative position (in most settlement cycles) under non-stressed circumstances depends on the type of clients (businesses or customers).

Figure 4.12 **Bank run client payments. Incoming and outgoing payments. September 2009 saw a temporary large outflow of client payments**



4.3.5.2 'Wholesale' bank run

If banks' customers, both companies and consumers, suspect a bank maybe about to fail, they can transfer their money to another account. The large-value payments of companies will become visible as client payments in the RTGS. The payments of consumers will become visible in the batch of a settlement organisation (see Section 4.3.5.1) or directly in case of urgent payments. These urgent payments are settled gross in TARGET2 within two hours after the instruction is given.

If customers have lost faith in their bank, they can massively send in client and urgent payment instructions and transfer their funds to another bank. In this case the same will happen as with a traditional bank run: the bank will soon run out of liquidity. The strength of this indicator compared to the cash bank run is that there is less fluctuation over time. This makes it easier to define a deviation from the normal patterns. Besides, banks might have sufficient cash available for the first few days of a bank run, which means that the bank run becomes only available a few days after it started. As a modified example, Figure 4.12 shows the in- and outflows of liquidity from client payments. A sudden sharp increase in outflow in this graph may signal customers losing faith. In this modified graph there was a strong outflow of liquidity in September 2009. After a few weeks this situation normalises again, which can be the result of eg state support (liquidity injection or nationalisation).

4.4 Behaviour of banks

4.4.1 Evidence from TARGET2-NL

As the payment system can be seen as a coordination game (McAndrews and Rajan, 2000, McAndrews and Potter, 2002 and Bech and Garratt, 2003), the participant's behaviour during the game is important. A study of the TARGET2-NL data of both the crisis period and the period before gives a lot of information about what is 'normal' and 'stressful' behaviour. A close look at what happened after Lehman Brothers' failure or before the nationalisation of Fortis and ABN AMRO, DSB Bank's failure and the state support, provides a wealth of information about the way banks react to shocks.

4.4.1.1 Changes in interbank loans

The interbank market after Lehman Brothers' failure decreased significantly. Evidence of this effect has been found by Heijmans et al (2010) for the Dutch market, Guggenheim et al (2010) for the Swiss market and Akram and Christophersen (2010) for the Norwegian market. Some banks in the Dutch payment system lend no liquidity anymore after Lehman's failure, even though they had sufficient liquidity available to do so. This is probably due to the market's increased perception of risk. Some banks which were generally a lending bank decreased their lending significantly in response to rumours and facts were known about these banks due to decrease in surplus. At the same time, when borrowing became more difficult for these banks,they had to have recourse to liquidity provided by the ECB (tenders and, in the end, marginal lending).

After the ECB had made year tenders available starting July 2009, many banks applied for these tenders as a security for potential future shocks. This liquidity was in many cases not necessary for the short term, as illustrated by the use of the overnight deposits directly after the tenders were received by the banks. The use of overnight deposits showed a peak just before the repayment the year after (July 2010) and gradually decreased to almost zero in March 2011.

4.4.1.2 Changes in timing of payments

Some banks that rumoured to suffer or actually facing liquidity and financial problems showed quire interesting payment patterns. We expected to see a slow increase in outgoing payments over time as a result of the worsened liquidity position. Initially this is exactly what happened. At first there was the expected delay in the payments', and also the average timing of incoming payments rose. However, the increase in average outgoing payments' timing was quickly followed by a sharp decrease to an average time below the 'normal' timing, with the average timing of incoming payments decreasing accordingly. This may reflect that the troubled banks wanted to give a clear signal to the market that they could meet their payment obligations in time. A troubled bank will continue this behaviour until it is no longer able to do this anymore for lack of sufficient liquidity.

4.4.1.3 Changes in collateral

In the heat of the crisis collateral was used much more often for intraday credit and monetary loans than before the crisis. Especially after Lehman Brothers' failure several banks faced difficulties funding themselves, which resulted in an increased use of collateral to keep fulfilling the payment obligations. The relative use of collateral increased for two reasons: 1) Some banks needed the collateral for other purposes. If these banks use the same amount of collateral for intraday credit and monetary policy, the relative amount increases 2) Other banks had to use more collateral to fulfil their obligations. In these cases also the absolute use of collateral for intraday credit and monetary policy increased.

4.4.1.4 Signs of a bank run

When the public became aware that a bank was having severe liquidity and financial problems and started worrying about a potential bankruptcy the data showed substantial outflows of banknotes, client payments and, in particular, emergency payments.

4.4.1.5 Existence of bilateral limits

It is expected that banks use bilateral limits to limit the maximum exposure to counterparties. Although, in TARGET2, banks have the possibility to use bilateral limits, this feature is not used by the banks participating through TARGET2-NL. This does not mean, however, that banks do not make use of bilateral limits in their own systems. The transactions between participants have been analysed for the existence of bilateral limits between participants. To find the limits in the data, we calculated the running bilateral net positions of two banks, A and B, during the day. When no limit was set by any of the banks, the running balances would be expected to follow a random walk. If, however, bank A had limited its position to a certain amount, the bilateral net position would often lie just beneath this limit, only to drop shortly as a result of incoming payments, and immediately to rise to this limit again as a result of outgoing payments . Bilateral limits will therefore deviate from random walks and thus immediately show up when time-weighted frequency counts of the bilateral balances are produced. The transaction data showed signs of the use of bilateral limits and even counter-limits by the other bank. When it was known that a bank was facing severe liquidity or financial problems, the bilateral limits for this bank were tightened to limit the exposure even further. From the discussions we have had with commercial banks, we also learned that banks grew more reluctant to allow a bank that might run into trouble to have a large negative position. This corroborates our expectation that such limits are applied.

4.4.2 Set of behavioural rules

Simulations are often based on historical transactions. The difficulty of using historical data is that these data usually do not reflect the stress scenario of interest. Therefore these historical data have to be modified to reflect this scenario. Part of this modification consists of introducing ‘adequate’ behaviour of banks in order to obtain realistic outcomes. The rules defined in this section are set up to be used for scenario analysis using historical data. Examples of scenarios are liquidity problems with a single bank, decreased trust in the interbank money market, changing monetary policy of the ECB, an operational outage of the payment system limiting the number of payments possible, etc. The scenario determines which of the rules have to be applied.

4.4.2.1 Preparation rule

Historical data not only contain regular payments, deposits of collateral and monetary loans, but also interbank loans, marginal lending and overnight deposit transactions. The latter are the result of a temporary shortage or surplus at the RTGS account. A shortage signifies that the cash reserve requirement cannot be met. A stress scenario for the payment system is designed to determine the new liquidity position , ie the liquidity shortage or surplus, if the scenario were to materialise. Therefore, when historical data are used the transaction data have to be cleaned for these funding transactions.

Preparation rule 1 *Historical transaction data used for scenario analysis of payment systems have to be cleaned for interbank loans, monetary policy transactions, marginal lending and ECB overnight deposit.*

4.4.2.2 Behavioural rules

The liquidity position of a bank can be influenced by all four elements in the overview of Figure 1: monetary policy, interbank market, collateral deposited and "real" payments. The behaviour (or policy) of the actor(s) behind these four elements determines what the effect on the liquidity position of these banks will be. We identify the following actors:

1. Monetary policy: the central bank
2. (Part of) the interbank market: banks that enter the market for lending and/or borrowing
3. Payments: banks and bank's customers (consumers and businesses)

4. Collateral: bank(s) depositing collateral for monetary and/or payment purposes. The central bank steers the eligibility and haircuts of the collateral, thus determining in the collateral value.

Monetary policy

The central bank uses its monetary policy to tighten or widen the money supply in the economy. If it wants to tighten the amount of liquidity, it can raise its main interest rate, the refinancing rate, making lending from the central bank more expensive. The more a bank lends or wants to lend the stronger the effect of a refinancing rate's increase will be for that particular bank. Another option central banks have, is to change the maintenance requirements. The higher/lower the requirement, the higher/lower the average amount of cash on banks' accounts need to be. This requirement affects individual banks differently as the requirement is bank-specific.

Behavioural rule 1 *Increase/decrease the access to tenders and/or decrease/increase cash reserve requirements depending on the central bank's role in the scenario.*

The Interbank market

The participants in the LVPS have several options to influence their counterparties' liquidity positions. The first option they have is to change the lending amount in the interbank money market. The level of trust a bank has in its counterparties determines the willingness to lend. If a bank does not trust (some or all of) its counterparties, it will decrease or cease its lending. If trust comes back, the lending amount will increase again.

Behavioural rule 2 *Decrease/increase the amount a bank can borrow in the interbank money market depending on the level of trust in this bank.*

Payments

Besides the interbank market, a bank can set bilateral and/or multilateral limits to one or more counterparties. A bilateral or multilateral limit is the most negative position this bank is willing to accept from a single counterparty or all counterparties respectively. If a limit is reached, the bank first needs to receive incoming payments either from the single participant (in case of bilateral limit) or one of the participants in the payment system (multilateral limits) before continuing to make its own payments. Section 4.4.1.5 described signs of bilateral limits' existence in our data. This can be translated into a behavioural rule by dividing the market into several groups:

1. Reliable banks: High rating and no rumours or negative news facts.
2. Less reliable banks: Lower rating or first rumours in the market.
3. Banks in trouble: Strong rumours and negative news facts regarding liquidity/financial problems.

Combining the three types of banks will lead to 9 possible outgoing payment flows with different setups of the bilateral limits, see Table 4.1 for examples of these limits. Banks with a high rating (bank type A) want to be seen as reliable by all counterparties. For other 'reliable' banks they will therefore observe a high bilateral limit. Regarding B-bank they will be slightly more reluctant but still accept a negative balance during the day. As regard to a bank which is the target of very negative publicity A-banks will be very careful and will only accept a relatively small negative intraday position. B-banks will observe slightly lower bilateral limits for B-banks and C-banks as their liquidity positions are somewhat worsened as a result of the first rumours in the market. Banks in trouble (C-banks) are basically compelled to change their limits to make sure they can pay as many counterparties. Due to the strict limits applying, this bank has to make sure it can have a negative balance with as many counterparties as possible in order to keep on receiving payments from its counterparties.

Behavioural rule 3 *Set bilateral limits depending on the type of bank.*

Table 4.1 **An example of bilateral limits of banks between the three groups**

	Receiving bank		
Sending bank	A	B	C
A	5% with max EUR 250 million	2.5% with max EUR 125 million	1% with max EUR 50 million
B	5% with max EUR 250 million	2% with max EUR 100 million	0.5% with max EUR 25 million
C	3% with max EUR 150 million	1% with max EUR 50 million	0.2% with max EUR 10 million

The percentages mentioned refer to the fraction of the total daily outgoing payment value.

Clients

At the moment when clients start losing faith in their bank there will be an increase in the outflow of liquidity. While the increase of payments as such cannot be controlled by the problem bank, the moment when these payments are settled can be (see below). Even though client payments are usually low relative to the total daily turnover of a bank's payments in the LVPS, the bank will increasingly be affected by the extra outflow and the liquidity position will worsen.

Behavioural rule 4 *Increase the outgoing payments' amount when the stress with respect to a bank continues.*

Bank in trouble

A bank facing liquidity problems cannot control the behaviour of its counterparties, central bank or its clients. It has however a few options to steer its liquidity position. We start with setting priorities to payments. Not all transactions in TARGET2 are equally important in terms of timing and impact. Continuous Linked Settlement (CLS) transactions eg have a very high priority and are very time-critical.[17] CLS is used by banks to settle foreign exchange transactions. The beneficiary of the transaction may be in another time zone and another

[17] CLS is a settlement process by which a number of the world's largest banks manage settlement of foreign exchange amongst themselves (and their customers and other third-parties).

LVPS. If this party does not receive the expected funds, it can face severe liquidity problems. The beneficiary commercial bank can either be another bank than the sender or a branch of the sending bank. The payments executed to settle the net balances of the EURO1 at the settlement account of the ECB and the payment obligations which result from settlement organisations (like eg the Dutch settlement organisation Equens) are also time-critical and therefore have a high priority.[18] It can be assumed that banks do not ignore or delay these time-critical payments intentionally.

Besides time-critical payments also payments on behalf of a customer will be executed relatively timely. Payment orders of customers with sufficient liquidity on their accounts will be executed within the normal time frame (of the contractual arrangement). Even though the bank might not have sufficient liquidity on its account, the customer does, and therefore the bank is obliged to meet its contractual obligations. Failure to do so might damage the bank's reputation and result in customer claims.

Due to the difference in time criticality the bank prioritises its outgoing payments. The most time-critical payments, like CLS, EURO1, and settlement organisations have the highest priority (priority 1). The client payments will have a slightly lower priority (priority 2). All other payments are considered less time-critical and therefore have the lowest priority (priority 3). If required, it is possible to define more levels of priorities.

Behavioural rule 5 *Transactions in payment system's scenario analysis have to be divided into priorities e.g.: 1) very time-critical, 2) time-critical and 3) other payment transactions.*

The second option a bank has to steer its liquidity position, is to change the timing of payments. In the transaction data of TARGET2-NL we have found that the troubled bank pays as soon as possible (see Section 4.4.1.2), which means as long as liquidity (balance and intraday credit) is available. If possible, it will start paying even earlier than it was used to do. If a bank were to delay its payments instead of paying in time and did so long enough, it would at some point not receive payments anymore as all bilateral limits to this bank would have been reached. Depending on the aim of the scenario, you

[18] EURO1 is a private sector owned payment system for domestic and cross-border single payments in euro between banks operating in the European Union. It is a net settlement system. Payments are processed throughout the day. Balances are settled at the end of the day via a settlement account at the European Central Bank.

let the bank in trouble pay as early as possible if you want to simulate a ‘natural’ reaction by the troubled bank. If you want to investigate whether delaying payments by a bank will lead to a gridlock, the payments should be delayed.

Behavioural rule 6 *Change the timing of the outgoing payments.*

Collateral

The last option a bank in trouble has to change its liquidity position is to change the amount of available collateral. As described in Section 4.4.1.3, our data showed that some banks in liquidity problems lowered their amount of collateral. Even tough this was a voluntary action by the bank in trouble, it had an adverse effect on their liquidity position. Ideally a bank in trouble brings in more collateral to increase its liquidity position (use liquidity for tenders and/or intraday) credit. Especially if the other actors (the market and clients) have changed their behaviour in such a way that it affects the liquidity position of the problem bank. Table 4.2 shows an example of collateral values banks have available for intraday credit for each of the three bank types, which can be used in scenario analysis. The central bank steers the eligibility and haircuts of collateral that is (to be) deposited.

Behavioural rule 7 *Decrease the collateral’s amount, which can be used for intraday credit and tenders, when the stress scenario aims to simulate severe problems with a bank.*

Behavioural rule 8 *Decrease the collateral amount, caused by reduced eligibility and/or increased haircuts of collateral.*

Table 4.2 **An example of collateral values of each bank type, which can be used for intraday credit**

Bank type	Collateral level sending bank
A	5 * daily average
B	2 * daily average
C	0.5 * daily average

4.5 Conclusions

This paper shows how the TARGET2 transaction and collateral data can be used to monitor banks. The monitoring looks at 1) the overall liquidity position, 2) demand and supply of liquidity, 3) timing of payments, 4) amount and use of collateral and finally 5) signs of a bank run. Combining the different elements of the monitoring gives information on the liquidity position of an individual bank. If just one of the indicator points in a certain direction, this need not mean much, but if more than one indicators point in the same direction it is clear that a bank is facing liquidity problems. If there are signs of liquidity or financial problems with a certain bank, this will first become visible in the ability to borrow the required liquidity. If the market perceives the risk as higher (or too high) that bank either has to pay higher interest rates for its loans due to this increased counterparty risk, or cannot borrow sufficient liquidity to fulfil its (maintenance) obligations. As a result of the lack of liquidity the troubled bank will proceed to make more intensive use of the ECB facilities (tenders and intraday credit) more intensively. Besides the inability to borrow the troubled bank faces a second liquidity decreasing measure by other banks by way of bilateral limit. These limits will reduce the negative position other banks are willing to accept towards the troubled bank. Even though these measures worsen its liquidity position, the troubled bank will seek to pay as early as possible to give a clear signal to the market that it is able to fulfil its obligations. By doing so it decreases the negative impact of the bilateral limits. The troubled bank is able to pay early as long as liquidity is available. If it runs out of liquidity it has no other option than to delay payments. If customers become aware of a bank's problems and expects a failure, they will either withdraw cash from the ATM or, more effectively transfer their money to another bank (either through client payments for the larger customers, or using urgent payments through internet banking). The liquidity position of the troubled bank then goes from bad to worse. Such a situation can usually only be solved with a market intervention (one or more banks take over this bank) or a state intervention (in terms of state support or nationalisation). If there is no intervention, the troubled bank will most likely collapse as the liquidity or financial problems are too big to be solved by the bank itself.

The set of behavioural rules described in this paper can be used in payment system stress scenario analyses, e.g. in simulations or network topology. The set of rules is based on the reaction patterns found in the TARGET2-NL transaction data and collateral management data before, during and after times of increased stress due to financial/liquidity problems of one or more banks. Using these rules improves the realism of stress scenarios and therefore the usefulness of its outcomes. The key features of the set of rules are divided into 1 preparation and 8 behavioural rules.

This paper shows that the TARGET2-NL transaction and collateral data gives valuable information on the liquidity position of banks. Close monitoring of banks using this data may reveal early signs of liquidity and/or financial problems. By looking into more detail it is possible to monitor the funding need and methodology of single banks, and how they change over time. After first identifying liquidity problems in the data or in other supervision information, the monitoring tool described in this paper can give up-to-date information on the problems' developments (up until the last business day if required). The information obtained from the TARGET2-NL transaction and collateral data renders it unnecessary to rely on the information given by banks themselves, which may be unreliable because facts may have been distorted or potential problems camouflaged. It is however not a substitute for current supervision information but a useful addition.

References

Abbink, K – Bosman, R – Heijmans, R – van Winden, F (2010) **Disruptions in large value payment systems: An experimental approach.** DNB Working Paper, 263.

Akram, F A – Christophersen, C (2010) **Interbank overnight rates – gains from systemic importance.** Norges Bank Working Paper, 11.

Allen, F – Gale, D (2000) **Financial contagion.** Journal of Political Economy, 108:1–33.

Bech, M – Atalay, E (2008) **The topology of the federal funds market.** Federal Reserve Bank of New York Staff Reports, 354.

Bech, M – Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory, 109:198–210.

Bech, M – Garratt, R (2006) **Illiquidity in the interbank payment system following wide-scale disruptions.** Federal Reserve Bank of New York Staff Reports, 239.

Bech, M – Soramäki, K (2005) **Systemic risk in a netting system revisited.** In Leinonen, H., editor, Liquidity, risks and speed in payment and settlement systems – a simulation approach. Proceedings from the Bank of Finland Payment and Settlement System Seminars 2005, pages 151–178.

Carlin, B – Lobo, M – Viswanathan, S (2007) **Episodic liquidity crisis: The effect of predatory and cooperative trading.** Journal of Finance, 62:2235–2274.

Cocco, J – Gomes, F – Martins, N (2009) **Lending relationships in the interbank market.** Journal of Financial Intermediation, 18:24–48.

Furfine, C (1999) **The microstructure of the federal funds market.** Financial Markets, Institutions and Instruments, 8:24–44.

Guggenheim, B – Kraenzlin, S – Schumacher, S (2010) **Solving the puzzel: Evidence of the unsecured Swiss franc money market.** Swiss National Bank Working Paper, 5.

Heijmans, R (2009) **Simulation in the Dutch interbank payment system: A sensitivity analysis.** In Leinonen, H., editor, Simulation analyses and stress testing of payment networks. Proceedings from the Bank of Finland Payment and Settlement System Seminars 2007–2008, pages 123–144. 28

Heijmans, R – Heuver, R – Walraven, D (2010) **Monitoring the unsecured interbank money market using TARGET2 data.** DNB Working Paper, 276.

Imakubo, K – Soejima, Y (2010) **The microstructure of Japan's interbank money market: Simulating contagion of intraday flow of funds using BoJ-NET payment data.** Monetary and Economic Studies, 28:151–180.

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 & Control, 32:259–278.

Koponen, R – Soramäki, K (2005) **Intraday liquidity needs in a modern interbank payment system: a simulation approach.** In Leinonen, H., editor, Liquidity, risks and speed in payment and settlement systems – a simulation approach., Proceedings from the Bank of Finland Payment and Settlement System Seminars 2005:71–114.

Lacker, J (2004) **Payment system disruptions and the federal reserve following September 11 2001.** Journal of Monetary Economics, 51(5):935–965.

Ledrut, E (2007) **Simulating relation in payment systems: Can banks control their exposure to a failing participant?** DNB Working Paper, 133.

McAndrews, J – Potter, S (2002) **Liquidity effects of September 11th.** Federal Reserve Bank of New York Economic Policy Review, 8:59–79.

McAndrews, J – Rajan, S (2000) **The timing and funding of Fedwire funds transfers.** Federal Reserve Bank of New York Economic Policy Review, 6:17–32.

Pröpper, M – van Lelyveld, I – Heijmans, R (2008) **Towards a network description of interbank payment flows.** DNB Working Paper, 177.

Shin, H (2009) **Reflections on Northern Rock: the bank run that heralded the global financial crisis.** Journal of Economic Perspective, 23:101–119.

Soramäki, K – Bech, M – Arnold, J – Glass, R – Beyeler, W (2007) **The topology of interbank payment flows.** Physica A, 379(1):317–333.

Wetherilt, A – Zimmerman, P – Soramäki, K (2010) **The Sterling unsecured loan market during 2006-2008: insights from network topology.** Bank of England Working Paper, 398. 2

Chapter 5

Combining liquidity usage and interest rates on overnight loans: an oversight indicator

Tatu Laine − Tuomas Nummelin − Heli Snellman

5 Combining liquidity usage and interest rates on overnight loans: an oversight indicator

Abstract

This study utilises payment system data to analyse market participants' liquidity usage and to trace interest rates paid for overnight loans. Our aim is to examine how liquidity usage has changed during the years 2006–2/2011 and to combine this information with data on overnight lending rates between market participants. It turns out that the Furfine algorithm used in the analysis produces overnight interest rates that correlate very closely with the EONIA curve. Based on the Finnish payment system data, we identify four separate time periods: normal, start of turmoil, acute crisis and stabilizing period. The results show that during the acute crisis period TARGET2 participants holding an account with the Bank of Finland paid on average lower overnight interest rates than other banks in the euro area. However, results reveal that there has been some lack of confidence between the Finnish participants since the onset of the financial crisis. A new indicator – the Grid – which we present here shows this very clearly. We suggest that this new indicator could be a highly useful tool for overseers to support the financial stability analysis.

5.1 Introduction

The payment system data include a wealth of information on market participants' behavior. The financial crisis has highlighted the need to better utilise also the data of payment systems in support of financial stability analysis. TARGET2 – the RTGS system owned and operated by the Eurosystem – provides real-time processing and settlement in central bank money. The data on payment transactions in TARGET2 could be used more effectively for oversight purposes, for instance to reveal potential problems of counterparties.

The purpose of this paper is to analyze the usage of liquidity by TARGET2 participants holding an account with the Bank of Finland. In addition, we analyze the interest rates paid by one financial institution to another for overnight credit. The aim is to combine these two pieces of information in building an indicator. Ideally, this indicator would be used in everyday oversight work, in order to immediately observe changes in a participant's behavior. Such an indicator would provide information from two independent sources, namely the central bank and individual participants in the interbank money market. The central bank can observe participants' accounts and liquidity usage in TARGET2, which information is otherwise known only to each participant. Participants see changes in other participants' behavior, if payment transactions are not received on time. However, banks make real time assumptions about each others' financial situation, as the interbank money market is highly integrated. If a participant assumes that another participant is running into trouble, the former is first supposed to raise the interest rate on overnight credit. If risks are assumed to be excessive, the participant is supposed to end its lending to the problem participant.

As stated above, this paper aims to combine earlier ideas into a single easy-to-use tool for everyday oversight work.[1] A further motivation is that interest rates on overnight loans have previously been studied using eg US and Dutch data but not Finnish data.

The results of our study indicate that the overnight loan interest rates combined with liquidity usage by market participants provides an indicator that could reveal whether a market participant is potentially in trouble. The crucial points here are that the data are analyzed on a daily basis and that the indicator is calibrated based on historical payment data. These are the next steps that should be taken.

5.2 Analysis

The analysis Section is divided into three parts. A brief literature review is given in 5.2.1. Part 5.2.2 describes how market participants' daily liquidity usage can be used to build the Forest Fire diagram. Part 5.2.3 describes how interest rates on overnight loans can be used to study the confidence of the market participants in each other.

[1] The basic idea was presented in the Bank of Finland Bulletin: Financial Stability 2010 (Bank of Finland, 2010).

5.2.1 Literature review

Liquidity usage has been illustrated by Forest Fire diagram in this paper. This idea was inspired by Capel et al. (2009) who presented the idea in the Bank of Finland Simulator Seminar in 2009. Furthermore, Heijmans and Heuver (2011) illustrated collateral use for intraday credit by using a similar diagram.

The other part of the indicator, namely interest rates paid on overnight credits, has been discussed in the literature of the last decade. Furfine (1999) developed an algorithm to trace transactions related to overnight loans and Demiralp et al. (2004) further developed the algorithm for selecting candidate loan transactions. Recently, Heijmans et al. (2010) improved the algorithm for identifying loans of maturities up to one year, in addition to overnight loans. Heijmans et al. (2010) show how spreads and volatility of interest rates on interbank loans increased during the financial crisis. Furthermore, Akram and Christophersen (2010) discussed overnight interest rates based on Norwegian data. They concluded that interest rates on overnight loans vary across banks and over time. In contrast, Eklund (2009) concluded that in Sweden the majority of overnight loans are made without a risk premium.

5.2.2 Liquidity usage

Liquidity management differs across participants in the TARGET2-Suomen Pankki system.[2] In this paper, liquidity refers to money in a participant's central bank account and to the collateralized overdraft facility of the account, which can be used immediately as intraday credit, when needed. The difference between the start-of-day balance and the minimum balance for the day is divided by the sum of the start-of-day balance and the available intraday credit, as shown below:

$$\text{Liquidity usage}\,[\%] = \frac{(\text{Start of Day Balance} - \text{Minimum Balance})}{(\text{Start of Day Balance} + \text{Intraday Credit})}$$

[2] TARGET2-Suomen Pankki system was launched on 18 February 2008, and it is part of TARGET2. We have used for the analysis the data of TARGET2-Suomen Pankki system and the data of its predecessor BoF-RTGS system.

This results in maximum usage of liquidity by the participant.[3] Some of the participants may actively manage their liquidity and extensively use intraday central bank credit. Some others may hold large deposits on their central bank accounts, in order to assure the smooth execution of payments and to avoid credit. High usage of liquidity, as opposed to active liquidity management, may also indicate problems for the participant in meeting its obligations.

Figure 5.1 illustrates liquidity usage of participants. The green area indicates participants using up to 30% of their liquidity and the black area indicates usage in excess of 90%. We see that in year 2006, ie before the financial crisis, more than one half of the participants used at most 30% of the liquidity. The green area shrunk after the start of 2007 and was at its smallest in 2009. The share of participants in the black area has been fairly stable, and less than 10% of participants have used over 90% of their liquidity. However, this share was slightly higher during 2007.

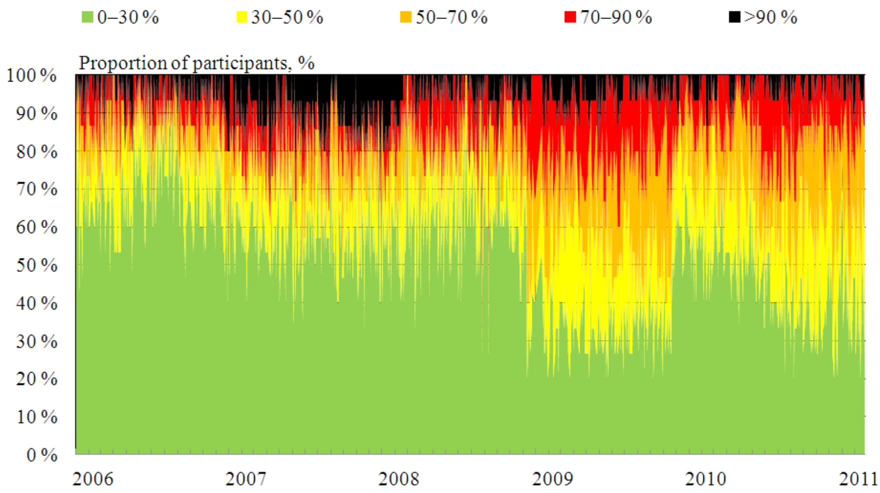
Problems in the US housing loan market escalated in August 2007 (at the start of turmoil period)[4]. In September 2008, US mortgage banks Fannie Mae and Freddie Mac were taken over by the federal government and Lehman Brothers went bankrupt (eg Bank of Finland, 2008). In this paper, this was defined as the start of the acute crisis period. As Figure 5.1 shows, the red, orange and yellow areas expanded after 2008Q3, which indicates that a bigger share of participants then used over 30% of their liquidity. Towards the end of 2009, the diagram shows that the liquidity usage is at the same level as during the start of turmoil period. However, since end-2010 the share of participants using less than 30% of their liquidity has been decreasing. This could indicate the forthcoming second turbulent period. Figure 5.1 shows the aggregate liquidity usage by participants, but for oversight purposes the liquidity usage of each participant could also be analyzed.

---

[3] The information on the value of additional eligible collateral pledged with the Bank of Finland could be added to the comparison, but these data were not available for this study.

[4] In this paper, we have divided the five years into four periods: A) the normal period before the financial crisis (1.1.2006–30.6.2007), B) the start of turmoil period (1.7.2007–14.9.2008), C) the acute crisis period (15.9.2008–30.6.2009) and D) the stabilizing period (1.7.2009–28.2.2011).

Figure 5.1 **Liquidity usage by TARGET2-Suomen Pankki participants**



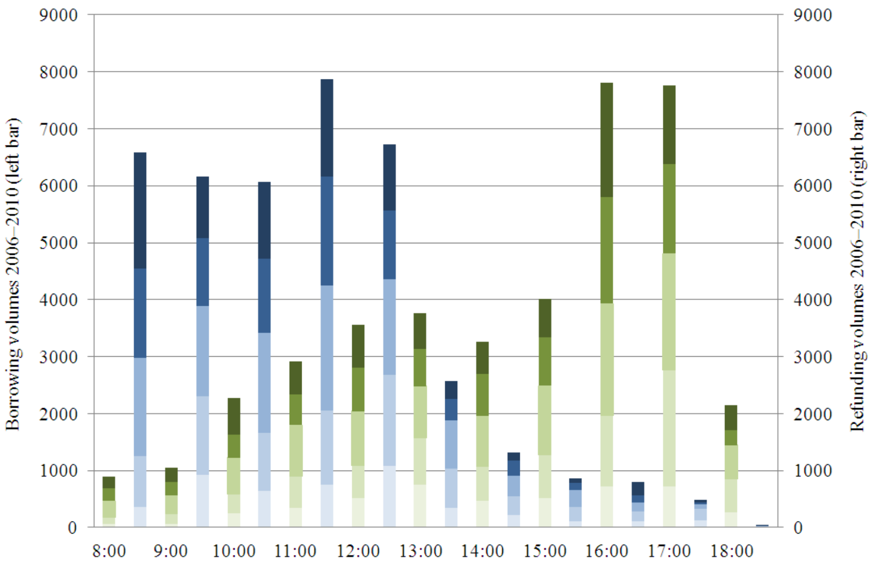
5.2.3 Interest rates on overnight loans

Normally, banks even out their liquidity fluctuations in the interbank money markets. If a bank pays a higher interest rate on overnight loans than others, this could indicate some lack of confidence. Furthermore, if there are notable changes in the level of interest rates paid by a bank compared to its own history, this could be an important signal, if interest rates paid by other counterparties remain steady. Information on interest rates on overnight credit is not readily available. However, interest rates can be estimated from payment systems data. We assume that loans and refunds take place in the same system.

In this study, we concentrate on overnight loans and do not analyze longer maturities. According to the results of Heijmans et al. (2010), over 80% of the value of interbank loan transactions was for overnight loans. Also in bilateral discussions with Finnish financial institutions, the participants have indicated that they mainly lend and borrow money from interbank markets on an overnight basis. Longer-term loans have remarkably decreased during the financial crisis. In this analysis, we concentrate on borrower-participants.

In bilateral discussions with financial institutions we have also found out the procedure for granting overnight loans (see Figure 5.2). Generally, the participants make deals before lunch time, after which the transactions are handled in the back-office and then executed in the payment system. Overnight loans of the previous day have to be paid back before granting new credit. The participants have a common understanding on the procedures of overnight loans. If a participant does not follow the unwritten rules, this may increase mistrust among other participants.

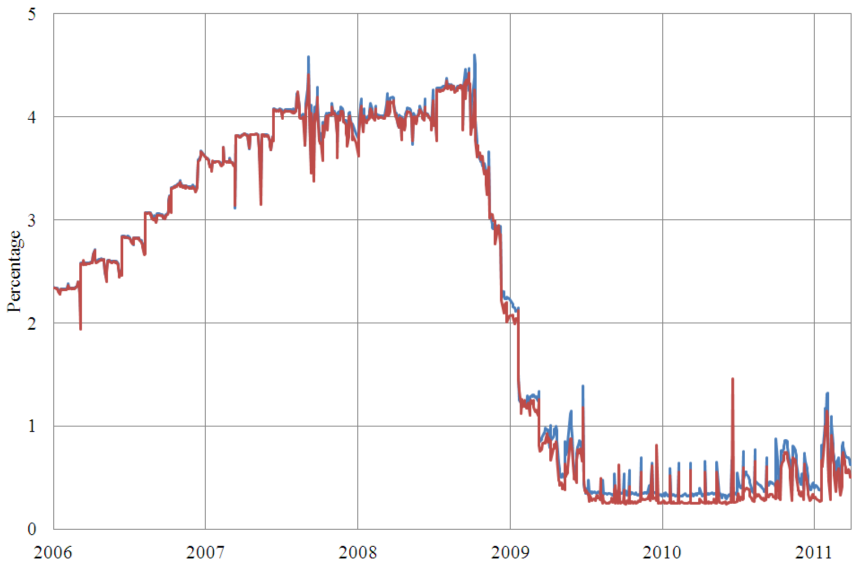
Figure 5.2 **Borrowing (left bars, green) and refunding (right bars, blue) numbers of overnight loans have different distributions over the business day. The light color (bottom of bar) indicates year 2006 and the dark color (top of bar) 2010.**



Following Furfine (1999), we search for transactions paid eg on Monday by bank A to bank B and refunded on Tuesday by bank B to bank A at interest rate r. The analysis is based on the actual data on TARGET2-Suomen Pankki transactions. Furfine (1999) used only payments larger than $1 million ending in five zeros whereas we included all transactions over EUR 10 000 ending in four zeros. Based on the discussions with market participants, the refund typically includes both principal and interest. Therefore, we assumed that loans are refunded on the next day and that these transactions include the original principal and the interest. In case of many matches, the first possible transaction pair was identified as the loan and refund (see eg

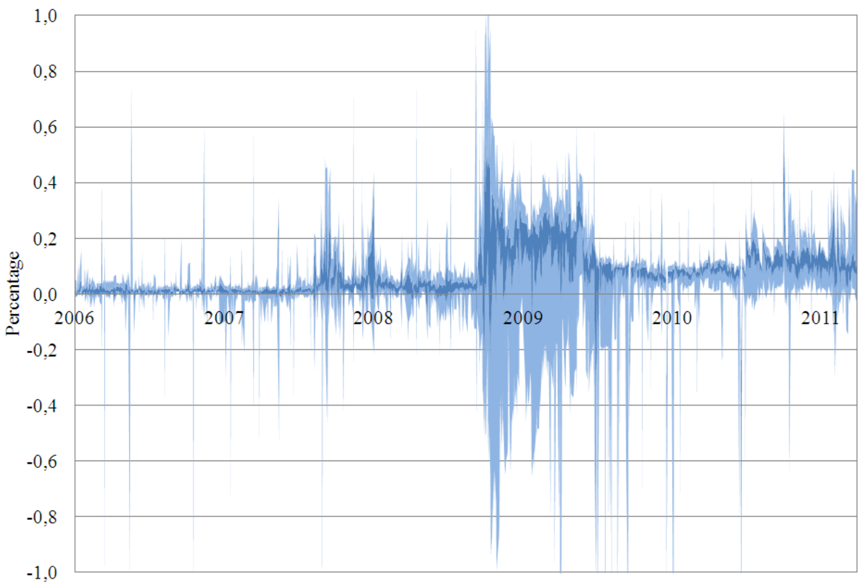
Heijmans et al. (2010)). The weighted average interest rate (FEONIA) was calculated from these found principal-interest rate pairs. Comparing FEONIA with EONIA indicates that these two interest rates follow each other very closely (Figure 5.3). In other words, the interest rate found on the basis of the Furfine (1999) method seems to work well. Quality of fitting can be described eg by the R^2 statistic, which is close to 1.

Figure 5.3 **FEONIA (Finnish EONIA; red line) and EONIA (Euro OverNight Index Average; blue line).**



During normal times, interest rates paid by counterparties seem to be highly concentrated and equal. Figure 5.4 is based on the daily data where interest rate paid by each participant was subtracted from EONIA. The spread between minimum and maximum values is shown, and from Figure 5.4 we can see that the interest rates paid by counterparties vary much more during the acute crisis period. This indicates that some counterparties have to pay higher interest rates because other counterparties have judged that the risks relating to those particular banks have increased. Figure 5.4 includes the data of participants in TARGET2-Suomen Pankki as well as the data on their transactions with other TARGET2 counterparties. If the RTGS accounts of foreign participants are excluded, the data describing the pure Finnish market are concentrated in the dark area. This means that banks participating in TARGET2-Suomen Pankki could have had overnight loans from the home market with lower interest rates than from abroad during the financial crisis. This indicates that domestic market participants had more confidence in each other than in foreign banks. It seems that TARGET2-Suomen Pankki participants pay lower interest rates on average than other TARGET2 counterparties after the acute crisis period.

Figure 5.4 **Spread between minimum and maximum of the difference curve (EONIA - interest rate paid by each participant for overnight credit). 50% of observations are concentrated in the dark area**



Single peaks were not smoothed out from the raw data of Figure 5.4. Most of the single peaks are false principle-interest findings[5] from the raw data, but if all of them are mechanically smoothed out, also the important change might not be observed. Here, particular days and participants were not reported, for the sake of anonymity. For oversight purposes, potentially stressed participants could be identified and monitored.

[5] Type 1 errors, ie some transactions are classified as overnight interbank loans even if they are not such transactions. See eg Heijmans et al. (2010).

5.3 Indicator

In this Section, we first present the general idea of the indicator. In part 5.3.2, we choose the two most interesting participants and conduct further analysis by including the time dimension. Further steps are discussed in part 5.3.3.

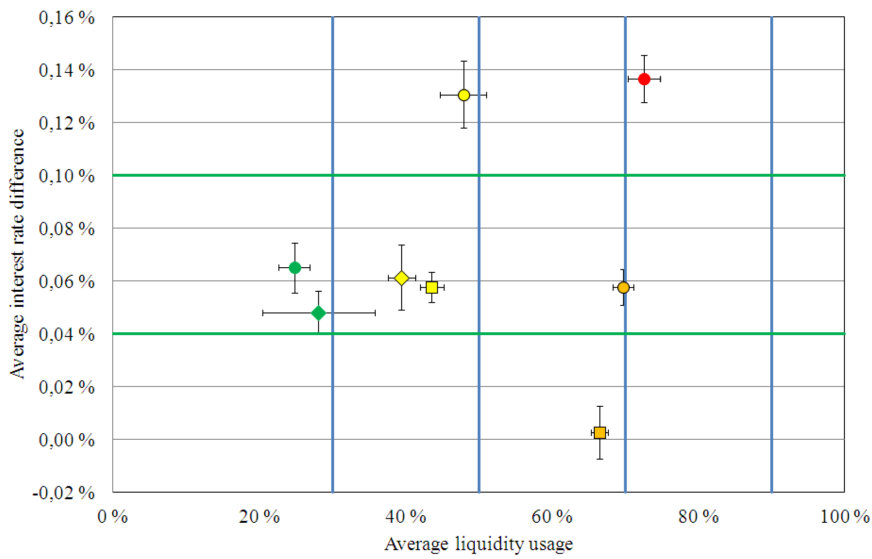
5.3.1 General indicator

The next step is to combine the information on liquidity usage and on interest rates on overnight credit. In more concrete terms, these two parts of the indicator can be included in the same plot; see Figure 5.5. In Figure 5.5, the average interest rate difference (EONIA − interest rate paid by each participant for overnight credit) calculated over about five calendar years (2006–2/2011) is shown as a function of the average liquidity usage for the eight biggest participants in TARGET2-Suomen Pankki system. From the figure we observe that the average liquidity usage varies from 25% to 73% and the average interest rate difference from 0.0028% to 0.1366%. The higher average interest rate difference indicates that the bank pays less for overnight credit.[6] From Figure 5.5 we see that the red dot (bank A) best manages its position. Bank A does not keep too much excess liquidity in the payment system and, on the other hand, it can get overnight loans at lower interest rate than the others. The orange square (bank B) has almost the same average liquidity usage level as Bank A, but it pays higher interest for overnight loans. Bank B could well deserve closer examination. The yellow dot (bank C) can also get cheap overnight loans, but its liquidity usage is not as efficient as for Bank A. The other banks stay in the neutral central zone. Their average liquidity usage varies from 25% to 67% and average interest rate difference from 0.0482% to 0.0651%.

[6] Generally, the Finnish overnight loan interest rates are lower than EONIA.

Figure 5.5 **A proposed indicator; the average interest rate difference is shown as a function of average liquidity usage.**

Now the total area is divided into fifteen blocks. A high interest rate difference combined with high liquidity usage indicates that a participant has a good and healthy position. A low interest rate difference combined with high liquidity usage indicates that a participant may be more stressed. Error bars indicate the area containing two-thirds of data points.



In the Finnish case, Figure 5.5 could be the baseline for the indicator. By defining the blue and green lines, we can divide the total area into subareas. In this example, blue lines define five liquidity slices (0–30%, 30–50%, 50–70%, 70–90%, 90–100%) and green lines three average interest rate difference slices (-0.02–0.04%, 0.04–0.10%, 0.10–0.16%). Altogether we have a grid of fifteen panels containing the participants. Since we do not have had any major bank defaulting in Finland, the large negative average interest rate difference values are missing. If a participant were in trouble, it would pay high interest on overnight loans and it would end up in the below -0.02% level in Figure 5.5. Also this same defaulting participant would have high average liquidity usage, probably above the 90% level.

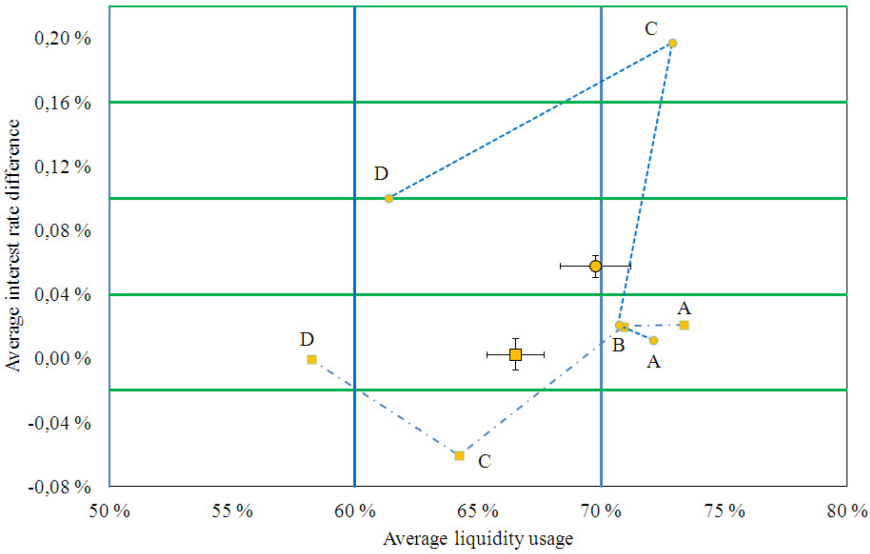
5.3.2 Time behavior

When building the general indicator in part 5.3.1 we found out that two participants should be further analyzed. We decided to divide the five years into four periods, as Heijmans et al. (2010) suggested in their work.

Figure 5.6 shows how the proposed indicator develops over time. The four points refer to time periods as follows: A) the normal period before the financial crisis (1.1.2006–30.6.2007), B) the start of turmoil period (1.7.2007–14.9.2008), C) the acute crisis period (15.9.2008–30.6.2009) and the D) the stabilizing period (1.7.2009–28.2.2011). From Figure 5.6 we see that two participants (dot and square) are very close to each other during the normal (A) and start of turmoil (B) periods. The participants move in opposite directions when the start of turmoil mode (B) turns to the acute crisis mode (C). Finally, the participants end up in different stabilizing period states (D). This means that the other money market participants have re-evaluated the creditability of these two participants. The dot participant is in a better position than the square participant after the turbulent period. The next step would be to monitor these two participants on a daily basis to see whether the disparity vanishes or continues to grow over time.

Figure 5.6 **An example of how the proposed indicator drifts over time**

The indicator is shown for two participants (dot and square) who move in opposite directions when the start of turmoil mode (B) turns to the acute crisis mode (C). The money market participants have greater confidence in participant dot than in participant square. (D) describes the stabilizing period state; the disparity of the two participants is clearly visible from the graph.



5.3.3 Further steps

To summarize, if the participant has to pay a high interest rate on overnight credit and, at the same time, has used almost all of its liquidity, this participant may be running into problems. Such observable change in behavior could guide an overseer or supervisory authority to take a closer look at the participant's behavior.

The crucial points are the calibrations of the indicator ie the Grid. In practice, this would mean studying in greater detail the historical payment data from 2006 to the present. The key question is which of the peaks in Figure 5.4 are true alarms and which are not. Part of the calibration procedure is how to choose time windows (eg daily, monthly, yearly average) for the data point in the Grid. The next step would be to bring the data processing more into real time. Each day, we should be able to calculate the indicator value on the basis of the previous day's data.

5.4 Conclusions

This work was inspired by the idea that information on TARGET2-Suomen Pankki payment system participants' liquidity usage could be combined with information on interest rates paid on their overnight loans. This Finnish example shows that there are observable differences between the ways in which participants manage their liquidity positions and how much they pay for overnight credit. Since there have been no major participant defaults in Finland, the results only indicate that some market participants do better than others, but that no one is in a serious trouble.

The next step could be to broaden the scope of this same exercise, eg to the 15–20 largest banks in the euro area. First, we should define the baseline period and block sizes for the Grid based on the historical data, which is not necessarily straightforward during abnormal times. The second step would be to detect and analyze the movements of participants within the Grid. The challenge here is to be able to calibrate the Grid smartly and to filter the single peaks, the false alarms, out of the raw data. If the filtering is too intensive, then there is a risk that also some significant sudden changes will be excluded. Obviously, Eurosystem non-standard monetary policy measures, like fixed-rate full-allotment procedures, have effect on the calibration of the Grid. In such a situation, participants may borrow more money from central banks, instead of interbank market.

Using the indicator also poses a challenge for the data management process, as data should be available on a close-to-real-time basis. In the best case, the today's indicator value would be calculated on the basis of yesterday's data.

In sum, we find this indicator to be a potentially highly useful tool for overseers to support the financial stability analysis. The more data from various sources are combined, the more knowledge we can obtain. Stock prices and collateral volumes could be next in line for adding to the indicator presented here.

References

Akram, Q F – Christophersen, C (2010) **Interbank overnight interest rates – gains from systemic importance.** Working Paper No. 11. Norges Bank.

Bank of Finland (2010) **Bank of Finland Bulletin: Financial Stability 2010.** Special issue, 37.

Bank of Finland (2008) **Financial Market Report.** 4/2008, 20.

Capel, J – Heijmans, R – Heuver, R (2009) **Behaviour of banks during the financial crisis.** Presentation in the 7th Simulator Seminar and Workshop 24–26 August 2009, arranged by Bank of Finland, see http://pss.bof.fi/Pages/Events.aspx.

Demiralp, S – Preslopsky, B – Whitesell, W (2004) **Overnight Interbank Loan Markets.** FEDS Working Paper No. 2004-29.

Eklund, J (2009) **Market structure, interest rates and the risk of financial contagion: An examination of the Swedish overnight market.** Master Thesis in Finance. Stockholm School of Economics.

Furfine, C (1999) **The Microstructure of the Federal Funds Market.** Financial Markets, Institutions & Instruments Vol. 8, No. 5, 24–44.

Heijmans, R – Heuver, R (2011) **Is this bank ill? The diagnosis of doctor TARGET2.** DNB Working Paper, No. 316.

Heijmans, R – Heuver, R – Walraven, D (2010) **Monitoring the unsecured interbank money market using TARGET2 data.** DNB Working Paper, No. 276.

Chapter 6

Monitoring the unsecured interbank money market using TARGET2 data

Ronald Heijmans – Richard Heuver* – Daniëlle Walraven*

* Heijmans and Heuver can be reached ronald.heijmans@dnb.nl and richard.heuver@dnb.nl, respectively. We thank Ron Berndsen, Hans Brits, Iman van Lelyveld and the referees of the Bank of Finland for providing useful comments The views expressed are those of the authors and do not necessarily represent the views of De Nederlandsche Bank.

6 Monitoring the unsecured interbank money market using TARGET2 data

Abstract

We investigate the euro unsecured interbank money market during the current financial crisis. To identify the loans traded in this market and settled in TARGET2, we extend the algorithm developed by Furfine (1999) and adapt it to the European interbank loan market with maturity up to one year. This chapter solves the problem of systematic errors which occur when only overnight loans are considered (as the Furfine algorithm does). These errors especially occur in times of (very) low interest rates. The algorithm allows us to track the actual interest rates rather than quoted interest rates on liquidity trading by participants of the Dutch part of the euro large value payment system (TARGET2-NL). The algorithm enables us to constitute the Dutch part of EONIA, making it possible to compare the interest rates developments in the Dutch market to the European ones.

6.1 Introduction

The interbank money market is important for banks and for the proper functioning of the financial system. Financial institutions sometimes face an expected or unexpected shortage or surplus on their accounts (Allen et al, 2009). They distinguish between two types of uncertainty concerning banks' liquidity needs. The idiosyncratic uncertainty arises from the fact that for any given level of aggregate demand for liquidity there is uncertainty about which banks will face that demand. The second type of risk concerns the aggregate uncertainty that is due to the fact that the overall level of the demand for liquidity that the banks face is stochastic. These (un)expected liquidity fluctuations not only affect the banks' liquidity management, but also impact the smooth operations of payments and RTGS systems (Iori et al, 2008). The interbank market is therefore an important element for an efficiently functioning financial system. In order to get a better insight into the Dutch part of the euro unsecured interbank market we use a methodology to identify interbank loans and their interest rates from TARGET2 transaction data.

This approach also enables us to monitor the functioning of this part of the unsecured interbank market during the recent crisis when liquidity trading was hampering.

In the unsecured interbank money market, banks with a shortage find banks with a surplus to trade the liquidity needed for doing their daily business. By trading in the interbank market they can fulfil their reserve requirements. Interbank market trading also provides an insurance against inter-temporal liquidity shocks (Bhattacharya and Gale, 1987). Allen and Gale (2000) show that liquidity shocks are the result of uncertainty in the timing of depositors' consumption. All these papers have in common that a well functioning interbank market is crucial for the ability of banks to access liquidity.

The majority of the loans traded in the unsecured interbank money market has a very short maturity, varying from overnight to one week. Banks are also trading liquidity with longer maturities but the longer the maturity, the more infrequent the trading. In the overnight interbank money market, EONIA is a benchmark interest rate.[1] The EONIA is an effective overnight rate computed as a weighted average of all overnight unsecured lending transactions in the interbank market, initiated within the euro area by the contributing panel banks.[2] The EONIA is quoted by this banking panel on a daily basis. Besides EONIA there is also EURIBOR, which is the rate at which interbank term deposits in euro are offered.[3]

The importance of the interbank market is well described in the literature. Cocco et al (2009) show that relationships are important for the ability to access interbank market liquidity. Due to the bilateral nature of this market, banks are able to establish such relationships. Apart from access to liquidity relationships do matter for both smaller and larger banks for receiving better terms both when borrowing and when lending (Cocco et al, 2009, Carlin et al, 2007). The model presented by Carlin et al (2007) indicates that under repeated interaction, cooperation among banks is an equilibrium outcome that involves refraining from predation and that allows those with a larger imbalance in their liquidity position to borrow at more favourable prices than they would otherwise. However, the effects of

[1] EONIA: Euro OverNight Index Average.

[2] The panel of contributing banks is 42 (September 2010), of which 3 banks connected to the Dutch part of TARGET2 (ING Bank, RBS N.V. and Rabobank), see http://www.EURIBOR-ebf.eu/.

[3] See http://www.EURIBOR-ebf.eu/.

relationships on pricing of liquidity differ among banks. Small banks tend to be net sellers of liquidity in the market but among all lenders, small banks receive lower interest rates than large banks for the funds that they lend. The larger the volatility the lower the interest rate that lenders receive on interbank loans (Cocco et al, 2009, Carlin et al, 2007). Large banks tend to be net borrowers in the market but they pay lower interest rates than small banks for the funds that they borrow. Borrowers with a higher proportion of non-performing loans tend to pay higher interest rates. The ability of larger banks to negotiate better lending and borrowing terms when lending to a counterparty with which they have a relationship is explained by their perceived too-big-to-fail image and their bargaining power (Cocco et al, 2009).

The fact that the interbank money market is crucial for the functioning of the banks themselves and the stability of the financial system was clearly illustrated by the current financial crisis in which banks became very reluctant to lend liquidity to each other. The current financial crisis, which started in the United States in the summer of 2007, not only thoroughly impacted the financial markets, it also showed the interdependence of the global financial systems.

To prevent banks from experiencing serious liquidity problems many central banks, including the European Central Bank, provided the market with additional liquidity. The ECB provided large amounts of liquidity, from mid-October 2008 on, through the policy of executing liquidity providing tenders at full allotment and a fixed rate. Moreover, the ECB introduced liquidity operations with longer maturities of up to 12 months and gradually lowered its target rate from 4.25% to 1%.[4] Banks used this extra liquidity to fulfil their funding needs, and/or to build up some liquidity buffers to withstand possible new shocks in this volatile and risk-averse market environment. These conventional and unconventional measures resulted in a high level of excess liquidity in the interbank money market. In addition to the decreased interest rates by the ECB, excess liquidity has put extra downward pressure on the overnight interest rate.

Despite the support from central banks, the situation in short-term unsecured euro interbank money markets is still fragile and the effects from the turmoil can still be felt. Turnover in the interbank money market has decreased significantly and trades are taking place at levels

[4] The first 12-month tender took place in June 2009. Another 12-month tender was executed in September and December 2009.

that are substantially lower than before the crisis. The lower trading volume can be explained by both increased risk-aversion among market participants and excess liquidity in the market. The higher risk-aversion is clearly illustrated by the extent to which banks made use of the standing facilities.[5] Institutions with a large surplus preferred putting their liquidity surpluses at the deposit facility of the ECB over trading it in the market, which would have been more profitable. After the bankruptcy of Lehman Brothers and the liquidity provision by the ECB many banks have put large amounts at the deposit facility of the central bank.[6] At the same time some banks had shortages and made use of the marginal lending facility of the central bank, while in normal conditions they could have borrowed the required funding in the market.[7]

The research question of this chapter is how the unsecured interbank loans can be identified from the TARGET2 transaction data. Furfine (1999) was the first to develop an algorithm to identify the interbank loans in (payment) transactions in Fedwire. Demiralp et al (2004) improved his algorithm to include more candidates for interbank loans. The Furfine and Demiralp et al algorithms have primarily been developed for the American interbank money market. The algorithm we develop in this chapter is suitable for the European unsecured interbank money market as the interest rates are based on the leading interest rates in this market: EONIA and EURIBOR. It improves Furfine's and Demiralp et al's algorithms as our algorithm can identify loans of maturities up to 1 year instead of only overnight. Ignoring maturities beyond overnight will lead to systematic errors in the interest rates found, especially in times with very low interest rates. Our algorithm reduces those systematic errors, and further more manages to describe the unsecured interbank money market more comprehensively. Because an accurate description of this market is important we need an algorithm that is functioning in both normal and stressed markets. In the investigated period, we have to deal with abnormal market circumstances as a consequence of the recent market turmoil. In order to give a better overview, we decided to split the investigated period into 4 different sub-periods

[5] The Eurosystem offers credit institutions two standing facilities: 1) marginal lending facility in order to obtain overnight liquidity from the central bank, against the presentation of sufficient eligible assets and 2) Deposit facility in order to make overnight deposits with the central bank.
[6] A maximum value of EUR 57.1 billion in 2008 and EUR 45.6 billion in 2009 in the Netherlands and EUR 89.5 billion in the first 6 months of 2010.
[7] With a maximum of EUR 1.6 billion in 2008 and EUR 4.6 billion in 2009 and EUR 0.1 billion in the first six months of 2010.

1. Pre-crisis period (01-01-2005 to 30-06-2007)
2. Start of the turmoil: (01-07-2007 to 14-09-2008)
3. Period after Lehman Brothers' failure and target rate changes by the ECB (15-09-2008 to 30-06-2009)
4. Start of unconventional monetary policy measures by the ECB (01-07-2009 to 31-10-2010)

The sub-periods in this chapter will be referred to as periods I, II, III and IV.

The outline of this chapter is straightforward. Section 6.2.4 describes the data set which is used for the analysis. Section 6.3 introduces the improved algorithm. Section 6.4 describes the difficulties with multiple matches. Section 6.5 discusses the developments in the Dutch interbank money market and section 6.6 contains conclusions and policy recommendations.

6.2 Large value payment systems

Large value payment systems (LVPS) play an important role in the financial system. Large value transactions, such as interbank loans, are mainly settled in these systems. Also the liquidity provided by central banks (including the ECB) as part of the monetary policy is done through the LVPS accounts.

6.2.1 Development of large value payment system

Central banks provide interbank systems that settle large value and time critical payments safely and efficiently. Over the past thirty years, turnover in interbank payments has increased enormously.[8] This increase is the result of financial innovation and the integration and globalisation of the financial sector. The settlement and contagion risk of netting systems were the reasons for many countries to develop Real-Time Gross settlement Systems (RTGS).[9] In RTGS systems each

[8] Development of the daily average value of the transactions of the Dutch market: EUR 10.3 billion in 1985, EUR 33.4 billion in 1995, EUR 83.1 billion in 2000, EUR 120.4 billion in 2005 and EUR 299.7 billion in 2010. The values of 2010 are partly due to some large foreign participants.

[9] Net settlement is a process in which transactions are not settled directly, but a total net position of all transactions is calculated and settled at the end of a given business cycle, traditionally a business day.

payment is executed immediately (real time) and individually (gross). The advantage of RTGS is that in case of default no payments have to be unwound. However, such systems require much more liquidity because for each payment sufficient liquidity has to be available on the account.

The large liquidity requirements of RTGS systems might lead to so-called grid-locks. These grid-locks occur when banks are waiting for incoming liquidity (payments) in order to be able to fulfil their own obligations. Central banks neutralise this problem by providing intraday credit to their banks. This intraday credit helps banks to execute their payments even if they lack sufficient liquidity on their account to do so. In most countries, including the European Union, this intraday credit must be collateralised and is free of charge. Intraday credit has to be repaid at the end of the day, lest it turns into an overnight loan for which an overnight fee is due. Simplified, in the United States banks are not required to supply collateral, but a fee has to be paid for the amount of intraday credit.

In 1985, three central banks implemented an RTGS system. By 1996, this number had increased to 16, mostly from industrialised countries. In 2006 the number of central banks with an implemented RTGS system was 93 out of 174 central banks (Bech and Hobijn, 2007).

6.2.2 The role of central banks

Central banks, co-ordinated by the Bank for International Settlement (BIS), have developed Core Principles, with which large value and other systemically important payment systems must comply (CPSS, 2001). Such payment systems must be safe and reliable to make sure that the probability of failure and abuse by others is minimised. Both market participants and central banks agreed that it is of the utmost importance to have quick settlement of transactions. This enables the receiving party to use the received liquidity immediately. Last but not least, a settled payment is final, meaning that a settled payment cannot be reversed, not even by a receiver in the event of failure. Received liquidity can thus be used to fulfil other obligations. Spindler and Summer (1994) state that there is an increasing need for central bank money, because this gives the best guarantee for the received liquidity (CPSS, 2003).

6.2.3 TARGET2

TARGET2, Trans-European Real Time Gross settlement Express Transfer, is the large value payment system of the Eurosystem, which is used to execute time-critical payments. Besides the euro countries, there are six non-euro European countries that are connected to TARGET2 for the settlement of euro payments.[10] Each central bank acts as the intermediary channel between a financial institution and TARGET2.

TARGET2 can only be used by institutions which meet the access criteria. The most important types of institutions that can gain access to TARGET2 are credit institutions established in the European Economic Area (EEA), national central banks of EU member states including the ECB and treasury departments of central or regional governments of member states active in the money market. Most other financial firms, non financial firms and consumers have no access to TARGET2.

6.2.4 Description of the data

The type of transactions we aim to identify in this chapter, the unsecured interbank loans, form a subset of the bank-bank and client payments but are unfortunately not labelled as such and cannot be identified easily. Most of the unsecured money market transactions are settled in TARGET2 and its predecessor TOP. Some banks also have the possibility to lend and borrow via the EURO1 system. However, we do not have transaction data of EURO1. In order to identify interbank loans we apply an indirect method described in section 6.3. The data set does not allow for selecting transactions that in fact are roll-overs or interest-only payments of derivative constructions.[11] We exclude these roll overs and interest-only payment transactions, because our algorithm does not permit them to be tracked and matched as there is no exchange of principal.

---

[10] Bulgaria, Denmark, Estonia, Latvia, Lithuania and Poland (status October 2010).

[11] A part of the interbank money market consists of transactions that are roll-overs and interest-only payments in case of derivatives constructions. In both cases there is a principal (loan or certain amount in derivative construction) over which interest is paid periodically. The principal amount is not exchanged between both parties but only exists in the contract.

The data used in our analysis consist of all bank-bank and customer-bank transactions of TARGET2-NL[12] and its predecessor TOP between 1 January 2005 and 30 April 2010. This excludes all Delivery versus Payments (DVP)[13] and lot settlement transactions.[14] In addition, the accounts of de Nederlandsche Bank and the Dutch Treasury (including its agency) are excluded as they are no commercial banks in TARGET2 (and TOP).

6.3 The 'Furfine' algorithm

6.3.1 The basics of the algorithm

Furfine (1999) was the first to create an algorithm to extract the overnight interbank loan transactions from the Fedwire payment system. This algorithm assumes a rounded value going from bank A to bank B at day t and the same value plus a plausible interest rate going in opposite direction at t+1. The interest rates (or in fact the values which are converted to a yearly interest rate) which he qualifies as plausible are based on the federal funds rate. Around this rate a so-called area of plausibility (or corridor) is defined within which the interest values must lie. Furfine chooses a corridor of 50 basis points (bp) below and above the federal funds rate. Stigum (1990) has argued that these loans have rounded values of one million dollars. Furfine uses this one million dollar as a minimum value for the loans and a round lot increment of 100,000 dollar. Demiralp et al (2004) use smaller sizes of loan-candidates of 50,000 dollar, and a round lot increment of 50,000 dollar. They note that an algorithm to filter interbank payments from a database could lead to Type 1 and Type 2 errors. A Type 1 error is a transaction mistakenly identified as an interbank loan while a Type 2 error is an interbank loan that is not found by the algorithm.

The reason for using an algorithm in the first place is that interbank loans settled in large value payment are not labelled as such in most systems (including TARGET2). It is also not known at what interest rates liquidity is traded in the market. Therefore assumptions have to be made to determine whether a transaction combination

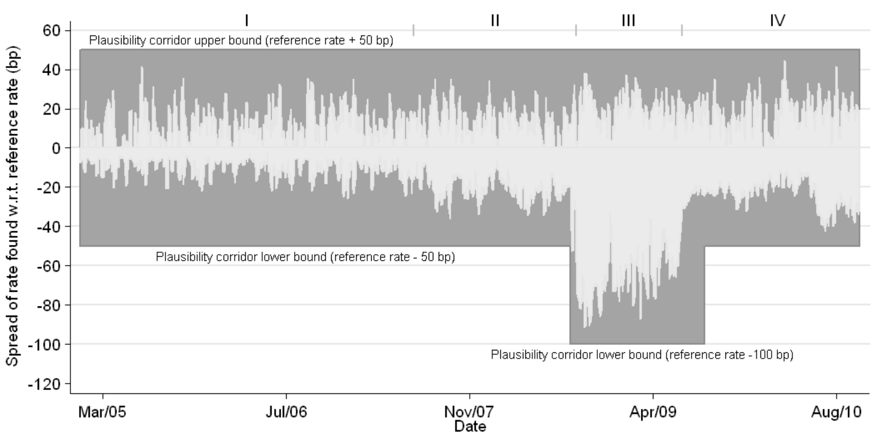
[12] TARGET2-NL was launched on 18 February 2008.
[13] DVP: The buyer's payment for securities is due at the time of delivery.
[14] The lot settlement transactions are the net settlement of retail payments by the settlement organisation Equens.

qualifies as a loan-refund match. The difference between the loan and refund value (the paid interest) can be converted to a year interest rate. As not all banks can trade liquidity at the same rates, an area of plausible interest rates has to be defined.

6.3.2 Areas of plausibility

The area of plausibility (or corridor) is chosen in such a way that it minimises the probability of Type 1 and 2 errors. Instead of the federal funds rate we use the leading European unsecured money market rates (EONIA for short maturities and EURIBOR for longer maturities) as input for the algorithm. We follow Furfine (1999) by using a corridor of 50 bp above and below the rates for most of the period. During the decrease of the ECB target rate and liquidity injection, it was necessary to increase the lower bound of the corridor to 100 bp, because the interest rates paid by some banks in TARGET2-NL (according to our algorithm) were below the lower part of the area of plausibility and would otherwise not be found (see Figure 6.1). This figure clearly indicates that from September 2008 to September 2009 an area of plausibility of 50 bp does not find all loans. Starting September 2009 the lower bound of the area of plausibility was set to -50 bp again, since a larger area was no longer necessary.

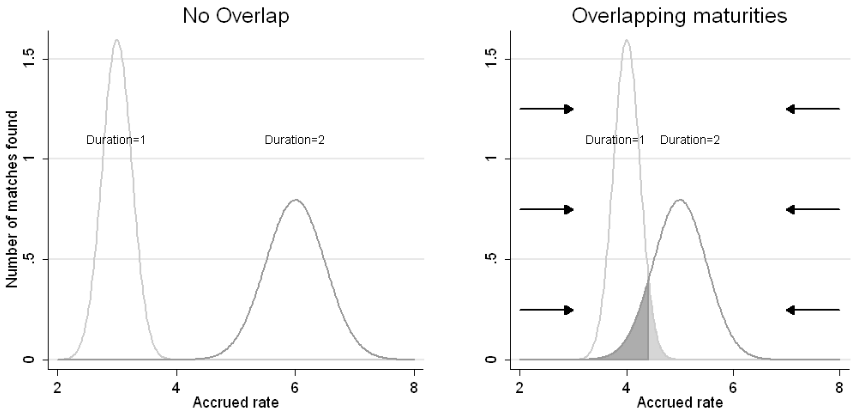
Figure 6.1 **Schematic overview areas of plausibility**



Another reason for choosing the corridor is the fact that we would like to simulate the interbank market structure to the largest extent possible. Under normal market conditions, EONIA is bounded by the official rates corridor fixed by the ECB (ECB, 2004). The upper bound is formed by the marginal lending rate at which banks can borrow against collateral. The floor of the corridor is set by the deposit rate at which banks can deposit funds overnight. In the period before the crisis, this corridor was 100 bp below and above the main refinancing rate, which was decreased to 50 bp in October 2008 and increased to 100 bp in the first quarter of 2009. From May until recently it was reduced to 75 bp in response to the crisis.

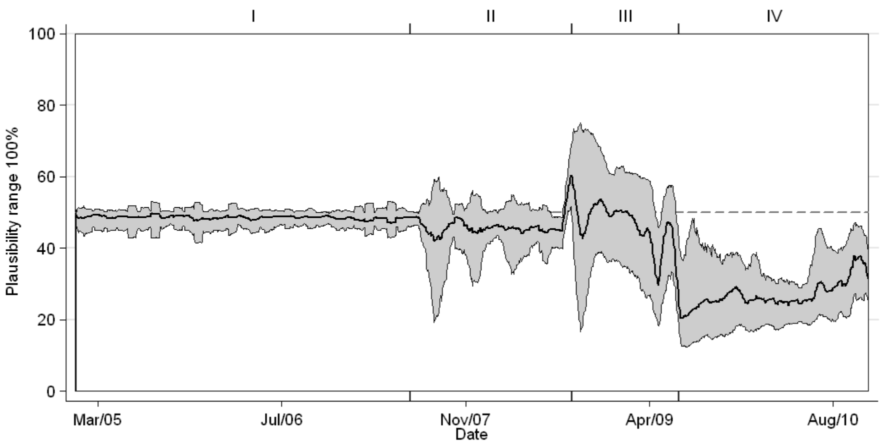
We started with a corridor of 100 bp below and above EONIA, since normally the main refinancing rate lies in the middle of the corridor and EONIA closely follows the official rate, mostly slightly above the main refinancing rate. By choosing these borders we expected to find the majority of trades and lowering the type 1 and 2 errors. Unfortunately, it turned out that this corridor caused too much overlap in maturities. Figure 6.2 gives a schematic overview of this overlap. The graph on the left of Figure 6.2 shows a situation in which there is no overlap. This graph shows two distributions of loans found by the algorithm for the overnight (duration 1) and 2 day (duration 2) loans. The two distributions do not overlap in this case. In other words, there is no uncertainty about the duration found by the algorithm. The graph on the right of Figure 6.2 shows a situation in which there is uncertainty about the maturity found. This graph shows the same two distributions as in the graph on the left but a certain area of the distribution belongs to both the overnight and two day loans. Loans found in this area can potentially be overnight or have a two day maturity. Besides the overlap problem the wide corridor might also lead to more Type 1 errors. With a corridor of 50 bp above and below EONIA, all our matches fit well within this corridor. We decided to use this corridor, because in the period before the crisis, banks trade liquidity against EONIA. Any deviation from EONIA is falling easily within the 100 bp total margin, so that the probability of type 1 and 2 errors are expected to be low. During the crisis, from October 2008 until May 2009, the ECB narrowed its corridor to 50 above and below the target rate, which makes our corridor structure even more similar to the actual interbank market structure. However, we are aware of the possibility that narrowing the corridor might lead to more Type 2 errors.

Figure 6.2 **Cross section of non-overlapping maturities (left hand panel) and overlapping maturities (right hand pane)**



Due to the lack of a database with traded interbank loans it is not possible to calculate the number of errors exactly. The way we control for this is to keep the corridor small, which leads to fewer Type 1 errors. But the corridor has to be large enough to prevent the algorithm from missing interbank loans (Type 2 error). To monitor the likelihood of the Type 2 error, we look at the distribution over the whole corridor of the overnight loan matches (see Figure 6.3). The black line represents the weighted average of interest rates on interbank loans as identified by our algorithm. The grey area represents 80% of the loans found (between the 10 and the 90 percentile). The figure shows that from January 2009 more interest rates are found closer to the boundary of the corridor. Banks at the boundaries of the plausibility corridor can borrow either very cheap or very expensive and have a larger probability of overlap between maturities, which is a clear signal that the probability of Type 2 error could increase. In the next section, we explain how we solved this problem.

Figure 6.3 **Distribution of the overnight matches within the plausibility area. Plausibility ranges stretched to 100%.**



6.4 Multiple matches

6.4.1 What are they?

6.4.1.1 Within the same day

There are several possibilities of matches between loan and repayment when using the algorithm as described above. The first combination of a loan and refund is a 1:1 match. This means a unique match as there is one loan with only one positive refund match. It is also possible that one loan has more than one positive refund match on one day or that more than one loan on a particular day has the same positive refund match, a so-called 1:N or M:1 match, respectively. The third possibility is that there are several loans on a particular day which have more than one possible refund match on another day, an M:N match. In case of an M:1 and a 1:N match, the first transaction is taken as the loan and refund, respectively. In case of an M:N match, the first loan and refund combination is taken, then the second and so on until there is either no loan or refund left to make a combination. The loan-refund combinations described in this paragraph are called intraday matches.

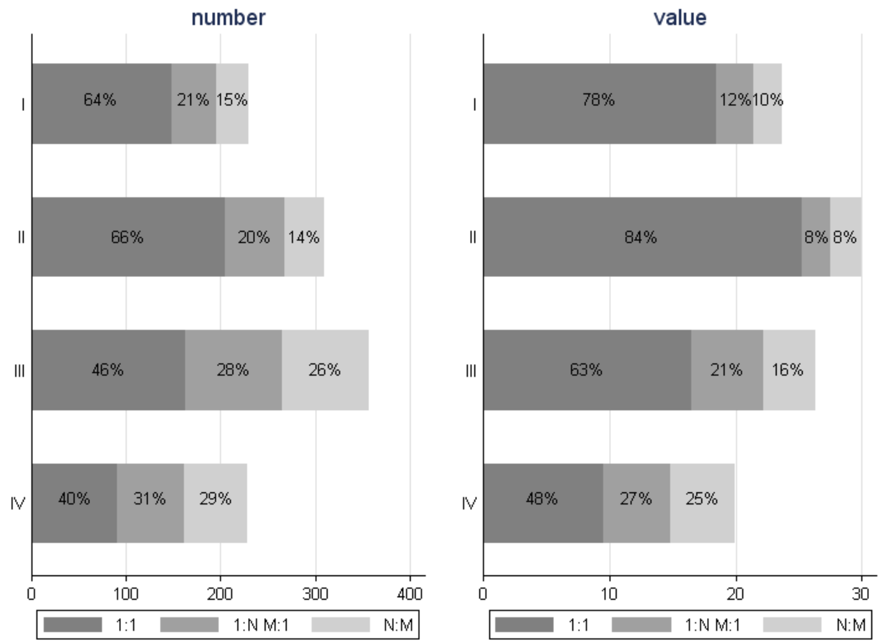
6.4.1.2 Between days

Matches also occur between different maturities. These inter-maturity matches occur if there is overlap between the corridors of two different maturities. The corridors of the overnight and 2-day loans do not overlap from January 2005 until approximately April 2009. Figure 6.2 shows what happens with the corridors of the overnight and 2 day two maturities. In an ideal world, there is no overlap between two maturities. Before the credit crunch, the overlap was relatively low, because the banking system was short of liquidity which brings interest rates up at a gradual pace with maturity. However, from April 2009 on interest rates are dropping and therefore the corridors of overnight and 2-day loans start to overlap (see the graph on the right of Figure 6.2). This means that an interest rate found within this overlapping region can be assigned both to overnight loans as well as two-day loans. The difficulties with the overlap is largely caused by the low interest rate environment and the fact that the market obtained long liquidity due to the ECB's conventional and unconventional monetary policy measures. These measures pushed the EONIA and EURIBOR rates down, and further increased the maturity overlap.

6.4.1.3 How many are there?

Figure 6.4 shows that, averaged over the four periods, we uniquely identify 57% of the number of transactions and 73% of the value. The variation in these percentages between those periods is large. During the fourth period, for instance, only 39% of the number of transactions and 48% of the value of the transactions are uniquely identified. The matching problem in the last period is the result of the market environment, which is characterised by low interest rates, over-liquidity in the market and the relatively large spread in interest rates between banks. In the second period, these percentages are 66% and 84%, respectively. In other words, the algorithm identifies the majority of transactions in a normally functioning market. However, in disturbed markets the algorithm encounters more difficulties but is still able to find around half of the interbank loans.

Figure 6.4 **Number and value of the interbank loans found per day by the algorithm divided into the four different periods. The multiple matches (1:N, M:1 and M:N) refer to the inter-maturity and intraday matches combined.**



6.4.1.4 How to minimise the inter-maturity matches

There are two approaches to reduce the problem with overlapping maturities. First, we could work with the interest rates that the algorithm found for each individual bank. Around this average bank interest rate, a smaller (less than 50 bp) corridor can be used. This solves the overlap problem by a) having a smaller corridor, which automatically leads to a smaller probability of overlap, and b) taking the corridor around the actual interest rates instead of EONIA. However, some caution is required using this method as it will only work for banks that are active and stay active in the interbank market. Some banks do not have (many) interbank loans, because they either fund themselves with liquidity provided by the ECB or are, in general, not very active in the interbank market. In these cases, no reliable average interest rates can be defined. For the large banks, however, this method will work well because they have been active in the interbank market during the whole period under investigation. We decided to solve the problem of overlapping maturities by letting the algorithm select the most plausible maturity, which is one day, full week(s) (1, 2, 3, etc weeks) and full month(s) (1, 2, 3, etc months). If one of the possible maturities is not one of the aforementioned, the shortest maturity is taken, because the market is deepest in the shorter end of the maturities.

6.4.1.5 Errors with the original Furfine algorithm

The algorithm developed by Furfine (1999) does not take the possibility of multiple matches into account, because it only addresses the overnight loans and therefore ignores the loans with a maturity longer than overnight. In period I, in 36% (22% in value) of the cases there is a non-unique inter-maturity match. However in period IV, the number of non-unique inter-maturity matches increases to 60% (52% in value). This is the result of the relatively low interest rates in this period and excess liquidity, which makes overlap between maturities more likely. The problem with ignoring inter-maturities overlapping matches is that it affects the interest rates of the found interbank loans. For example, a 2-day loan could be wrongly marked as an overnight loan resulting in higher interest rates found for the overnight loans.

6.4.2 Our algorithm

Our algorithm searches for interbank loans and refunds using the following criteria

1. The loan must be a rounded value larger or equal to EUR 100,000 (increment EUR 100,000).
2. The refund must be equal to the loan plus a plausible (positive) interest rate

$$X_{t+d} = X_t + \delta(d) \tag{6.1}$$

   With X_t the initial loan value, X_{t+d} the repayment of the loan with duration d (in days), $\delta(d)$ the plausible interest rate dependent on the duration.

3. Euro trades are based on 360 days,[15] which is relevant for calculating the year interest rate.
4. The plausible 'central interest rate' ($i_{central}$) used, is EONIA or EURIBOR. The EONIA rate is applied in accumulation for maturities of up to 4 days. For maturities of 5 days and longer, the corresponding EURIBOR rate is applicable.

$$i_{central} = \text{EONIA if } 0 \le t \le 4 \text{ days} \tag{6.2}$$

$$i_{central} = \text{EURIBOR if } T \ge 5 \text{ days} \tag{6.3}$$

5. A match will occur when the interest rate found (i_{found}) falls within a corridor of 50 bp below and 50 bp above $i_{central}$. The lower bound of the corridor is set at 100 bp below $i_{central}$ from September 2008 to September 2009. The absolute minimum lower bound is set at 5 bp. The δ in the formula under point 2 has to follow the following criteria

$$(i_{central} - 0{,}5\%) \cdot X_t \le \delta \le (i_{central} + 0{,}5\%) \cdot X_t \tag{6.4}$$

$$(i_{central} - 1{,}0\%) \cdot X_t \le \delta \le (i_{central} + 0{,}5\%) \cdot X_t \text{ for September 2008} - \text{September 2009} \tag{6.5}$$

$$0{,}05\% \cdot X_t \le \delta \tag{6.6}$$

6. The loan and the refund of the interbank loan are equally routed. In other words, both transactions are processed between the same two BIC-codes and within the same system, which is either TARGET2 or EURO1. In our model, it is only possible to identify a loan that is made fully within TARGET2.[16]

Even though the algorithm described in this section may still have some imperfections, it gives a reliable impression of the interbank money market. We focus entirely on the unsecured part of the interbank money market, because the lack of information regarding collateral streams prohibits us from focussing on both secured and unsecured money markets.

---

[15] In the United Kingdom, trade is based on 365 days.

[16] It has been confirmed by liquidity managers of commercial banks in the Netherlands that it is common practice that loans and reimbursement take place in the same system.

6.4.3 Testing our algorithm

In order to check whether the algorithm reflects the unsecured interbank money market accurately, it is compared with a real market environment. The trading platform e-MID is chosen for this check, because it is a part of the euro unsecured interbank market environment and the trades executed at the e-MID platform, like most loans within the Eurozone, are settled in TARGET2.[17] The algorithm's viability and reliability in reflecting the interbank market has been checked by correlating the EONIA according to the model (Dutch EONIA) with the rates at which liquidity is traded at the e-MID platform during the same period. The daily (value weighted) average of interest rates found by our algorithm, Dutch EONIA, correlates highly with the interest rates applied to trading at the e-MID-platform (the R^2 is almost 1).

We also checked for the presence of seasonal effects. Seasonal effects due to reserve requirements and accounting reasons (such as window dressing) at the end of quarters, half year and year have been observed, but do not have any disturbing effects in our algorithm because it is part of a natural market functioning. Moreover, these effects are sufficiently small so that the chosen area of plausibility does not have to be adjusted.

6.5 Results

6.5.1 Developments in the interbank market during the crisis

Our algorithm enables us to investigate the specific characteristics of the Dutch part of the unsecured interbank market structure and its changes as a consequence of the recent credit crunch. First of all, we would like to know at which interest rates banks in TARGET2-NL actually borrow and lend liquidity. We call this rate the Dutch EONIA or DEONIA. The DEONIA is the effective overnight rate of all overnight unsecured lending transactions that have actually taken place in the Dutch part of the euro interbank unsecured money market.

[17] E-MID trades deposits denominated in four currencies and is owned by 29 banks and the Italian Banking Association. Market participants from 26 countries are active on this trading platform. There are 4 large banks active on the e-MID platform, which participate in TARGET2-NL.

The DEONIA is not computed using market quotes but real transactions. Having insights into actual borrowing and lending rates gives us more information about the functioning of this market and its individual participants.

Figure 6.5 presents the development of the spread of DEONIA to EONIA during the period under investigation. Before Lehman Brothers' failure, the DEONIA follows EONIA very closely. Interestingly, the figure indicates that the more stressful the market is the more the DEONIA diverges from EONIA. In other words, banks in TARGET2-NL are able to attract unsecured liquidity at lower interest rates than EONIA.

In period I, most banks could lend liquidity at rates which were close to EONIA. The average daily standard deviation in the interest rates in period I was 4.1 bp. Looking at variations among types of banks (Figure 6.6), we observe that larger banks were able to lend liquidity at interest rates very close to or slightly above EONIA (their spreads fluctuate around -0.3 bp).[18] At the same time, they were able to borrow at more profitable rates, indicated by a downward deviation of the prices they paid compared to EONIA (the average spread is -1.3 bp). In lending liquidity, smaller banks follow the same pattern as their larger peers. However, they lend (average spread of -0.7 bp) and borrow (average spread of -1.9 bp) their liquidity at more profitable rates than larger banks. Foreign banks have to pay the highest price for liquidity (average spread is -1.9 bp), but are able to borrow at even more favourable terms than the large Dutch banks (average spread is -2.4 bp). These results for the Dutch part of the unsecured interbank market are consistent with the finding in the literature that larger banks are able to negotiate more favourable lending and borrowing terms than smaller and foreign banks (Cocco et al, 2009, Freixas and Holthausen, 2005). Nevertheless, we observed that foreign banks manage to negotiate the best borrowing terms. We will examine this in more detail later in the chapter.

We observe a change when markets became more stressed, starting in the summer of 2007 and reaching its top in October 2008. The more stressful the market, the higher the variability in interest rates at which banks borrow and lend liquidity are. This is consistent with the

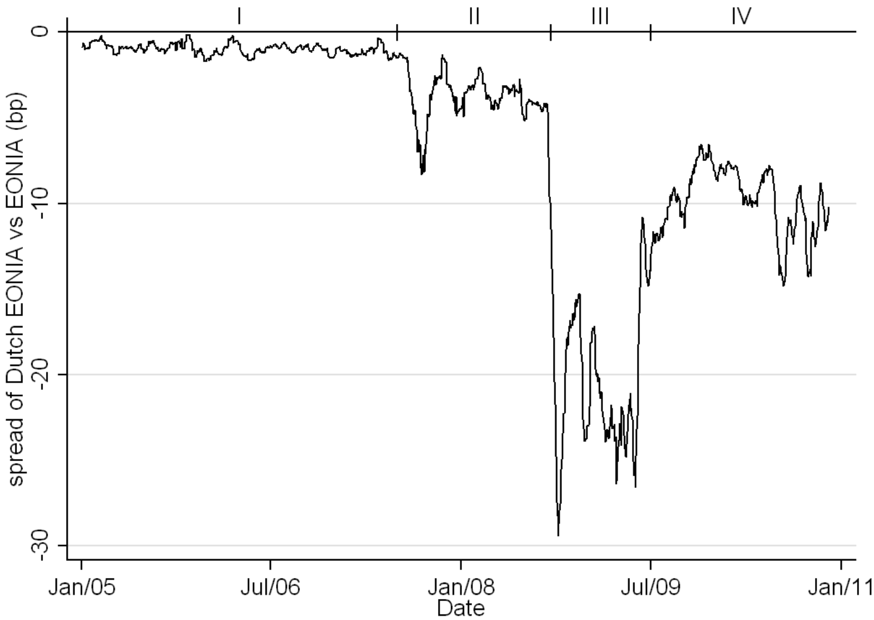
[18] Some caution is required comparing the group of foreign banks in TOP and TARGET2 because this group is partly different. Especially some large British banks are participating in TARGET2-NL, which were not a participant in TOP. Besides some of the foreign banks were only a small participant in TOP and a big participant in TARGET2-NL. The activity of those banks on the interbank market is larger in TARGET2 than in TOP.

argument of Gorton (2009) that the more the crisis unfolds, the larger are the volatility and spread in rates at which liquidity is borrowed and lent due to higher risk aversion in the market. Moreover, we observe that this effect is strongest in the lending side of the market which can be explained by the higher exposure to counterparty risk. From October 2008 the spread in DEONIA developed from 7.1 bp for period II, to 18.6 bp for period III and 10.1 bp for period IV.

We investigate whether the credit crunch impacted the borrowing and lending rates for different types of banks. We find that it contributes to a stronger differentiation among banks in rates at which they can borrow and lend liquidity. Contrary to the ‘normal’ market situation, smaller Dutch banks are now able to lend at higher interest rates (Figure 6.6) than the larger banks. This is in contrast to the expectation that large banks are able to get more profitable lending conditions, even in stressed markets due to their market power and too-big-too-fail character (Cocco et al, 2009).

Foreign banks were lending liquidity at the lowest interest rates compared to their large and small Dutch peers. Our results suggest that foreign banks still had access to liquidity but at more unfavourable conditions than before the Lehman collapse. This is consistent with the literature (Freixas and Holthausen, 2005) that suggests that foreign banks could have more difficulties in trading money in stressed markets. Foreign banks are lending at interest rates that are even below the ECB’s overnight deposit rate. This is an interesting phenomenon, because normally the ECB’s deposit rate forms the lower bound of the rates at which liquidity is traded in the unsecured interbank market. Apart from the potential higher counterparty risk perception, the observed situation can also be explained by the fact that these market participants do not have access to the standing facilities of the ECB. It could be more profitable for these banks to lend their excess liquidity abroad, eg in the Dutch market, at rates below the ECB’s deposit rate than to deposit their money at their central banks’ overnight deposit facility at even lower or no remuneration. This is consistent with the findings of Akram and Christophersen (2010) who study the overnight interbank interest rates paid by banks in Norway.

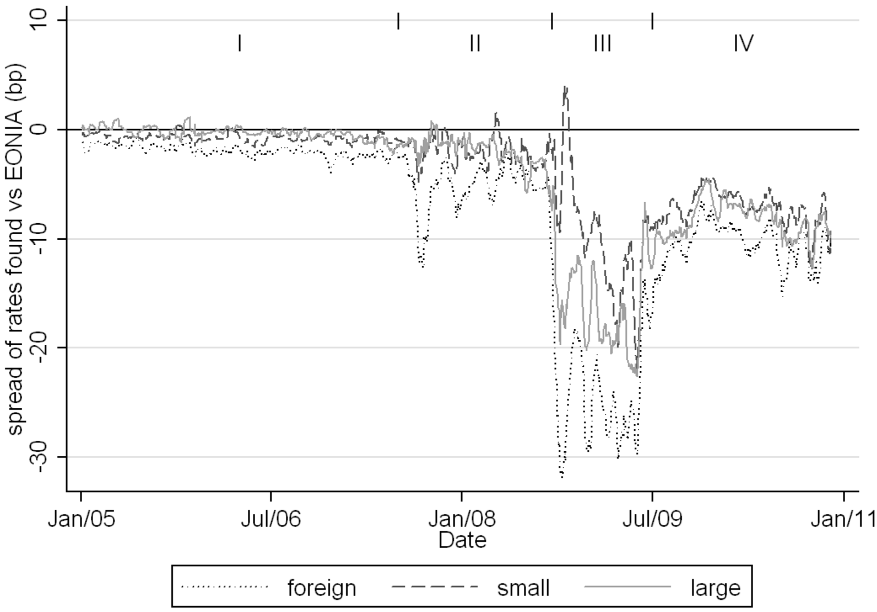
Figure 6.5 **Lending spread of Dutch EONIA to EONIA**



On the borrowing side, variability in prices of liquidity traded increased when market stress rose. Also the price variations increased for different types of banks. But compared to the lending side, tensions were slightly lower on this side of the market. Before market tensions started, banks could borrow at rates below EONIA. Smaller banks borrowed at higher rates than large banks and foreign banks. Foreign banks borrowed at lowest levels, but this situation changed in the summer of 2007. From then on, larger banks negotiated the best borrowing rates. In the heat of the crisis, foreign banks paid the highest rates apart from a short period of improvement in the first quarter of 2009 (February and May 2009). In the summer of 2009, the variability in borrowing rates among the different types of banks slightly decreased. Foreign banks, however, had to pay the highest rates again. Large banks borrowed money at similar rates. Smaller banks, in contrast, were able to borrow money at lower rates than large and foreign banks.

In the recently stressed, unsecured interbank market it appears that smaller banks are lending and borrowing at the best terms. This is in contrast to expectations that larger banks are negotiating the best lending and borrowing terms due to their too-big-too fail character and bargaining power (Cocco et al, 2009). The results show that these advantages can fade away even for larger banks if markets become more stressed and uncertainty and risk-aversion increase. Uncertainty about the exposure of large Dutch banks to US subprime loans reduced the willingness of other market participants to trade with these peers and induced them to look for counterparties that seemed to have fewer difficulties, like smaller banks because of their lower perceived subprime exposures.

Figure 6.6 **Lending spread of bank groups' Dutch EONIA to EONIA**



Result 1 *There is a significant increase in the spread and volatility of interest rates at which banks in TARGET2-NL lend and borrow after the credit crunch hit the market.*

Result 2 *There is a significant increase in the spread and volatility of interest rates at which liquidity is traded among different types of banks (small, large and foreign) after the credit crunch.*

Result 3 S*maller banks are able to negotiate the best lending and borrowing conditions during the crisis while the large banks had the best conditions before the credit crunch affected this market.*

Result 4 *Foreign banks lend money at rates which are even lower than the ECB's deposit rate, but foreign banks borrow at highest rates in the Dutch market. This combination of high borrowing rates and low lending rates implies that these banks faced least favourable conditions during the credit crunch.*

Further investigation of the unsecured Dutch interbank money market makes clear that the majority of the interbank transactions in both value and number have a very short maturity. Almost 59% of the total number of loans and almost 90% of the total value of the loans has a maturity of 1 week or shorter. Fifty percent of the number of transactions (with maturities up to 3 months) is overnight. The second most frequently traded loans are weekly loans, with a share of 4.4%. Besides the overnight and 1 week loans, liquidity is also traded regularly in the duration buckets of 2, 3, 4, 5 etc weeks and 1, 2 and 3, etc months. The overnight loans are most important with a share of 82% of the total value, followed by the one week loans with a share of 4.4%.

Result 5 *The majority of the transactions in the unsecured Dutch part of the money market take place in the very short end: 50% of the number of transactions and 82% of the value are overnight.*

A normal functioning unsecured Dutch interbank market shows some periodically higher trading activity due to reserve requirements and financial reporting. At the end of the reserve maintenance periods, banks trade more heavily in the interbank market to fulfil their reserve requirements. A similar reasoning explains the higher market activity at the end of months, quarters, 6 months and year as a consequence of accounting smoothing. These behavioural aspects are typical for the interbank money market and are also visible in other markets (Iori et al, 2008, Cocco et al, 2009).

Result 6 *There is an increase in the interbank loan activity at the end of the reserve maintenance period and the ends of (financial) reporting periods.*

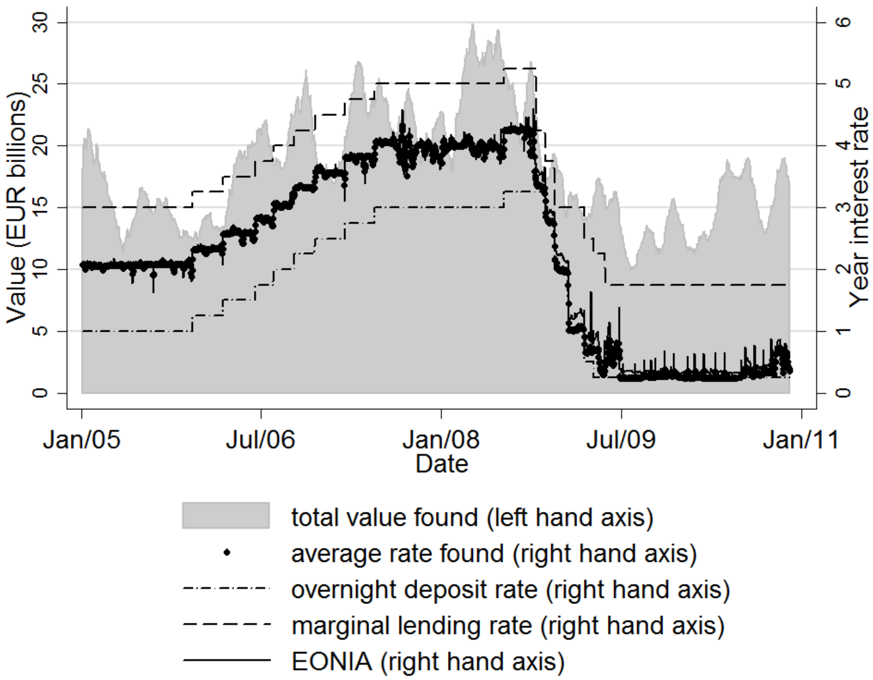
Our data-analysis confirms the significant decrease in the number and value of the overnight loans traded in the Dutch part of the euro unsecured money market, since the start of the crisis in September 2008. The average total turnover decreased from EUR 23.5 billion in period II to EUR 17.3 billion in period III and EUR 13.4 billion in period IV (see Figure 6.7). However, it seems that there has been some relief in the stressed market in the summer of 2009, because there is an upward trend in unsecured overnight lending. In this period, the ECB started injecting extra liquidity into the system by its special, very long-term tender operations (duration of 1 year) and other unconventional monetary policy measures.

As a consequence of the increased risk-aversion during the credit crisis, we expected to see a clear shift (from longer) to shorter term loans after the failure of Lehman Brothers. This shift is, however, not supported by our findings. The euro money market survey (ECB, 2009) and anecdotal evidence suggest that this could be explained by market participants who are changing their trading activity from the unsecured to the secured interbank market, especially in the short end. The preference for secured lending lies in the fact that the collateral that is part of the secured trading reduces (counterparty) risk. We have to rely on this evidence because we are not able to check this in our model, because of a lack of secured money market transaction data.

Result 7 *The credit crunch resulted in a significant decrease in the volume of unsecured liquidity traded in the Dutch part of the euro unsecured interbank. But the market did not completely vanish. Contrary to our expectations, there was no clear shift from long(er) to short(er) term unsecured lending.*

Result 8 *The overnight lending activity in de Dutch part of the euro unsecured interbank market increased slightly since July 2009, when the ECB started its unconventional monetary policy activity.*

Figure 6.7 **Total amounts and interest rate of the overnight interbank loans in TARGET-NL for the period 01-01-2005 to 30-06-2010. The total value reported is a 23-day moving average.**



6.5.2 Monitoring

A profound insight into interbank money markets is crucial for maintaining financial stability and executing monetary policy. The value added of our algorithm can be found in the information it provides about the interest rates and volumes at which banks actually trade in the unsecured Dutch part of the euro unsecured interbank money market. The developments in this market can be monitored at

1. the macro level: this level looks at the behaviour of all (or a large group of) banks combined.
2. the micro level: this level zooms in on individual (or a small group of) bank(s).

The translation of the algorithm in a monitoring framework can be used as an early warning system for sudden shocks in the market or (slowly) worsening of market conditions. The latter refers to the macro level as well as the micro level monitoring. It is to be expected that in many cases certain events will be visible at both levels. Section 6.5.3 describes the monitoring at the macro level and section 6.5.4 describes the monitoring at the micro level.

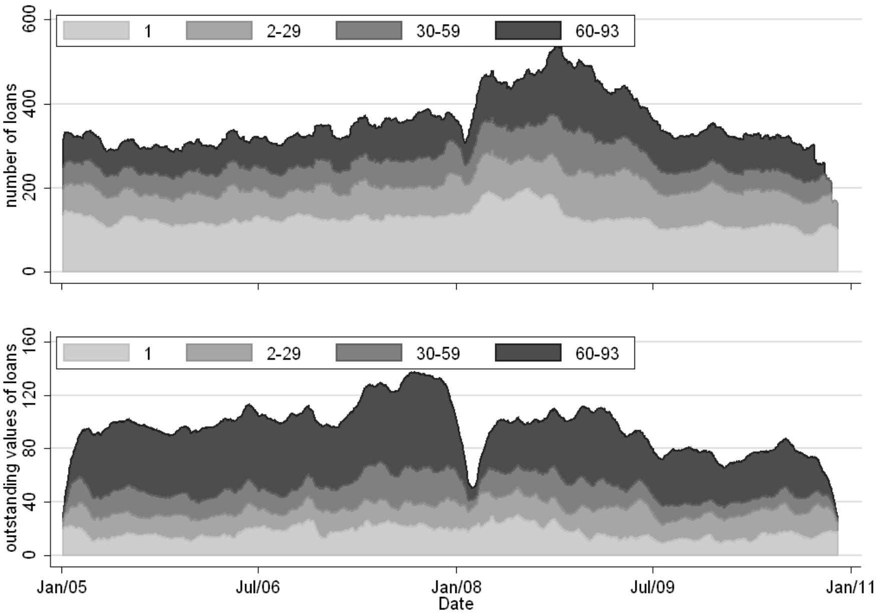
6.5.3 Macro level

The macro level gives an overview of the state of the interbank money market as a whole and will be particularly useful in discovering market trends. By tracking developments on a macro level, changes in market structure can be observed in an early stage. The ability to monitor developments in the unsecured Dutch money market at a regular (even daily) basis is particularly interesting because the effects of monetary policy decisions can be tracked continuously. In this respect, it is very interesting to monitor the effects of exiting unconventional monetary policy measures by the ECB in the near future.

In monitoring the interbank market, the EONIA curve is used as a benchmark for developments in the European interbank money market (see Figure 6.7). Against this curve, the Dutch Interbank Money Market rate or Dutch EONIA is plotted (red dots). The Dutch EONIA is made up of the rates at which banks in the Dutch part of the interbank money market have traded. With these two curves, it is not only possible to examine the trends in the European and the Dutch interbank money markets, but also to monitor the spread. Moreover, it is possible to plot developments in the value and volume of the loans settled in TARGET2-NL.

The graph in the upper part of Figure 6.8 shows the number of loans for each day in the sample period. Some caution is required with the changeover from TOP/TARGET to TARGET2 and the last 3 months of the investigated period (light blue parts). In the changeover period to TARGET2 it is not possible to identify cross border loans which started before the migration date and were refunded after this date. Moreover, for the last 3 months of the period under investigation, it is not possible to identify all the loans yet, because the loans can only be identified by matching them to a refund. No loans with maturities beyond 1 September 2010 can be identified, because the refund has not been paid yet. Therefore the last three months show partial results for the longer maturities.

Figure 6.8 **The top graph shows the number of loans that is traded in every maturity bucket. The bottom graph does the same for the outstanding value.**



In the change over period from TARGET tot TARGET2 and in the last three month of both graphs not all durations are available (yet). The numbers 1, 2-29, 30-59 and 60-93 in both legends refer to the maturities included in the representation.

The graph in the lower part of Figure 6.8 shows the outstanding total value of the loans including maturities from overnight up to 3 months. An interesting aspect is that roughly 50 percent of the outstanding value has a maturity of longer than 1 month.

Figure 6.5 shows the spread in the interest rates found by our algorithm (= Dutch EONIA) and EONIA. A strong deviation from EONIA is visible in the period when the ECB lowered its target rate from 4,25 percent to 1,00 percent. In the last year of the sample period, the interest rates in TARGET2-NL are still lower than EONIA, even though EONIA is historically low in this period with values varying from 29.5 to 69.0 bp.

6.5.4 Micro level

The micro level analysis provides insight into individual lending and borrowing behaviour. The single bank focus is of utmost importance, because specific market trends cannot only affect banks differently but can also be the beginning of a development that will impact the unsecured money market at the macro level. The micro level monitoring tool also provides detailed information on the performance of individual banks participating in TARGET2-NL. These data can be compared with common market developments. Similar graphs as for the macro level can be developed for each bank. Relevant parts of the unsecured money market for an individual bank are the lending and borrowing rates and volumes. The rates of lending and borrowing should also be seen relative to the market part (e.g. the Dutch market) or a group of banks (eg high or low-rated banks).

6.6 Conclusions

The algorithm developed in this chapter has extended the algorithm developed by Furfine (1999) and Demiralp et al (2004) by making it applicable for the euro unsecured interbank money market, with a specific focus on the Dutch part of this market. Our algorithm makes it possible to investigate the unsecured interbank money market more extensively, because the leading interest rates (EONIA and EURIBOR) are part of our algorithm and we are able to identify transactions with maturities of up to 1 year. Furfine and Demiralp et al. focused primarily on overnight loans. Moreover, the algorithm provides information about the interest rates at which banks have traded in the euro interbank unsecured money market. This information is valuable, because it provides more insight into developments and possible distortions in the market than the quotes making up EONIA. Lastly, our model has been translated into a macro and micro level monitoring tool for this market.

The results of our algorithm can be used as a very useful monitoring tool of the unsecured euro-denominated interbank money market. The high correlation of the average interest rate found by our algorithm, both with e-MID rates and EONIA, proves that the algorithm functions well (the Type 1 and 2 errors are relatively small). The outcome of our algorithm shows that ignoring loans with longer maturities can lead to large systematic errors, especially when the reference interest rate (in our case EONIA) is very low. In particular, identification of maturities longer than 1 month is still a challenge if the plausibility corridors become quite large, which results in an increase in overlap between two or more maturities.

Monitoring of the market as a whole works very well using the results of our algorithm. However, some caution is required for the results for individual banks, especially those at the boundaries of the plausibility corridor, because some banks can borrow either very cheaply or very expensively. Matches at the boundary of the corridor have a larger chance of overlap between maturities. For banks active in the interbank market the problem of overlapping maturities may be reduced by using an individual bank interest rate as reference rate and a smaller plausibility corridor around this interest rate ($i_{central}$). In case banks are not active in the market, the problem can be solved by proceeding from the shortest maturity

As expected, the total value of the interbank loans in the unsecured money market has decreased significantly to approximately 50 percent of the value traded in the period before Lehman Brothers' failure. The overnight market accounts for 50 percent of the number of transactions and 82 percent of the value of these transactions over the four periods. The shift from long to short-term lending, however, is not supported by our algorithm. The Dutch EONIA and EONIA are very similar before the credit crunch, but the Dutch EONIA deviates from EONIA when market stress increases. There is a clear increase in the volatility of interest rates compared to the pre-crisis period. In other words, the perceived increased counterparty and liquidity risk for some banks is also visible in the interest rates of Dutch banks. A clear difference between individual banks is also visible when the credit crunch intensified, as interest rates at which participants were trading became more volatile and the spread of rates at which different types of banks borrow and lend liquidity increased. The ease at which bank have access to liquidity has become more dependent on individual bank characteristics. Smaller banks were able to negotiate the best lending and borrowing conditions during the crisis contrary to large banks, which had the best terms before the credit crunch. Foreign banks had the least favourable conditions during the credit crunch, because they borrowed liquidity at highest rates and lend money at rates below the ECB's deposit rate. The latter is unexpected and an interesting phenomenon because the ECB deposit rate is regarded as the lower bound in the market. However, this situation could exist because at the time interest rates were very close to the overnight deposit rate as a result of the low interest rate environment and liquidity in the market.

Using the results from the algorithm developed in this chapter, it is possible to obtain a good overview of the unsecured euro-denominated interbank money market. The different aspects of the crisis are clearly visible in the interbank market.Monitoring the interbank market at both macro and micro level gives a good understanding of the current status of the money market. The monitoring indicators could also be used for policy making regarding the unsecured interbank money market and monetary policy. After the failure of Lehman Brothers, the ECB introduced unconventional monetary policy measures. It is a generally accepted idea that the liquidity provision of central banks at the current rate and amount cannot continue forever. When central banks are executing their exit strategies, the monitoring tool described in this chapter is very valuable for tracking the effects at the macro and micro level. Especially the monitoring at the micro level will be interesting, because individual banks will, most likely, not react similarly to the upcoming policy measures of central banks. This is partly caused by the segmentation of the market into groups of banks on the basis of differences in perceived counterparty risk. In short, this tool allows us to monitor the stability of the unsecured Dutch interbank money market.

References

Akram, F A – Christophersen, C (2010) **Interbank overnight rates – gains from systemic importance.** Norges Bank Working Paper, 11.

Allen, F – Carletti, D – Gale, D (2009) **Interbank market liquidity and central bank intervention.** Journal of Monetary Economics, 56:639–652.

Allen, F – Gale, D (2000) **Financial contagion.** Journal of Political Economy, 108:1–33.

Bech, M – Hobijn, B (2007) **Technology diffusion within central banking: The case of real-time gross settlement.** International Journal of Central Banking, 3:147–182.

Bhattacharya, S – Gale, D (1987) **Preference shocks, liquidity and central bank policy.** Cambridge University Press, Cambridge.

Carlin, B – Lobo, M – Viswanathan, S (2007) **Episodic liquidity crisis: The effect of predatory and cooperative trading.** Journal of Finance, 62:2235–2274.

Cocco, J – Gomes, F – Martins, N (2009) **Lending relationships in the interbank market.** Journal of Financial Intermediation, 18:24–48.

CPSS (2001) **Core Principles for systemically important payment systems.** Bank for International Settlements.

CPSS (2003) **The role of central bank money in payment systems.** Bank for International Settlements.

Demiralp, S – Preslopsky, B – Whitesell, W (2004) **Overnight interbank loans.** Manuscript Board of Governors of the Federal Reserve.

Freixas, X – Holthausen, C (2005) **Interbank market integration under assymetric information.** Review of Financial Studies, 18:458–490.

Furfine, C (1999) **The microstructure of the federal funds market.** Financial Markets, Institutions and Instruments, 8:24–44.

Gorton, G (2009) **Information, liquidity and the (ongoing) panic of 2007.** American Economic Review, 99:567–572.

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 & Control, 32:259–278.

Spindler, A – Summer, M (1994) **The central bank and the payment system.** Washington DC: IMF.

Chapter 7

Evaluating the impact of shocks to the supply of overnight unsecured money market funds on the TARGET2-Banca d'Italia functioning: a simulation study

*Luca Arciero**

* Luca Arciero. Banca d'Italia. Market and Payment Systems Oversight Department. Email to Luca.arciero@bancaditalia.it. The opinions expressed herein are those of author and do not commit the Banca d'Italia. I would like to thank Gronqvist, C; Impenna, C and Marseglia, G for useful comments at different stage of this work. I wish also thank the participants at the 7th Simulator Seminar at the Bank of Finland held in Helsinki in 2009 and those at the CPSS Workshop on Payment System Monitoring Indicators held in Basel in 2011. I would also like to thank Harry Leinonen, Matti Hellqvist and colleagues at the Bank of Finland for making available the BoF-PSS2 Simulator. The remaining errors are author's responsibility only.

7 Evaluating the impact of shocks to the supply of overnight unsecured money market funds on the TARGET2-Banca d'Italia functioning: a simulation study

Abstract

This paper presents a simulation exercise assessing the ability of Italian banks to fulfil their payment commitments in TARGET2, the Euro area Real Time Gross Settlement System, in the event of a contraction in the supply of funds in the overnight unsecured money market. The results of the exercise, which was carried out with reference to two reserve maintenance periods: 12 November to 9 December 2008, during the acute phase of the crisis, and 11 November to 7 December 2009, show that even a drastic reduction in trading on the interbank market would have caused only limited effects on the system's functioning. These broadly positive results depend basically on two factors. First, both simulations refer to periods in which the ratio of unsecured overnight loans to the total number of settled payments was significantly lower than in the pre-crisis period; and in both periods the average level of liquidity in participants' settlement accounts was relatively high due to banks' reliance on Eurosystem credit facilities.

7.1 Introduction

Large value interbank payment systems are used by banks to settle obligations stemming from their own activity in financial markets and transactions on behalf of their customers. These systems are a pivotal part of the financial system architecture, and their smooth functioning is crucial for the conduct of monetary policy and maintenance of financial stability.

The majority of large-value interbank payment systems use real-time gross settlement (RTGS) as the modality for settling payments. Because payments are settled individually upon entry into the system, provided the sender has sufficient liquidity in its account, an RTGS system requires large amounts of liquidity to enable participants to smoothly process their payment obligations.

Deposits borrowed from other banks in the money market – especially those negotiated at the overnight maturity – represent one of four possible sources of funds for an individual participant to meet its intraday payment obligations in such a system; the others comprising balances held in central bank accounts, incoming payments from other banks, and (if available) daylight credit facilities provided by the central bank.[1]

System participants' reliance on the money market as a source of liquidity makes the RTGS system and the money market deeply interrelated, but, although the relationship between money market trading volume and RTGS systems performance has been widely discussed in both the theoretical (Angelini, 1998, Bech and Garrat, 2002) and policy (BIS, 1997) literature, little empirical effort has been devoted to this issue so far.

Among the existent empirical literature, it is worth recalling the contribution by Aschraft et al (2007) who explain the intraday allocation and pricing of overnight loans of federal funds in the form of settlement balances held by dealers in the US RTGS system (Fedwire); and the 'agent based' model developed by Arciero et al (2008), which features the main elements of a real-life RTGS system – a central bank acting as liquidity provider and a simplified money market – and shows that the money market may play a fundamental role in the evolution of system functioning in the wake of a critical event. Finally, Klee (2010) analyses how Fedwire-participants' operational problems in submitting payments affect the behavior of the federal fund rate, showing that deviations of the federal funds rate from the policy rate (ie the Federal Open Market Committee's target rate) are related to the severity of the operational outage, the payment volume sent by the affected participant, and the time of day at which an outage occurs.

The scarcity of empirical contributions is largely attributable to the endogeneity that emerges in the relationship between flows exchanged in the money market and the performance of the settlement system. When facing the adverse effects of a shock in the money market, banks' desks may modify either the mix of funding sources (eg by

---

[1] Money market funds are a peculiar funding source as they '… can only serve to redistribute funds already within the system, although that may nevertheless make an important contribution to reducing the reliance on banks' reserve balances and central bank credit extensions…' BIS (1997).

augmenting intraday credit lines at the central bank) or the management of payments (eg by changing the timing of payment submission by reducing their arbitrage-oriented money market trades). Moreover, the endogeneity may be is strengthened if the central bank decides to inject additional liquidity into the system when money market conditions deteriorate.

The difficulty of finding a valid instrument to overcome this endogeneity suggests the need for a scenario-based study: in a simulation exercise, it is possible to run a series of counterfactual scenarios by replicating a number of operational days in the life of an RTGS system under varying money market conditions, holding constant the behavior of central and commercial banks.

The aim of this paper is twofold: to empirically study the efficacy of the money market – versus other liquidity sources – for the smooth functioning of an RTGS system, and (for policy purposes) to gain insight into the resiliency of TARGET2-Banca d'Italia to a contraction of the money market more severe than those experienced during the financial turmoil.

To achieve this aim, we rely on the widely used *Bank of Finland Payment and Settlement System Simulator* (BoF-PSS2)[2] and, although the money market comprises a large set of instruments and maturities, we focus on: i) the overnight maturity[3] – because it is arguably the more relevant source for intraday payment purposes; ii) the unsecured segment only – because, as witnessed by the last crisis, it is more likely to dry up in a situation of financial stress.

In this vein, four sets of simulations were run, mimicking two different money market shocks, a shock to cash balances, and a shock to daylight credit lines.

The first set of simulations was aimed at evaluating the ability of Italian banks to fulfill their payment commitments in the event of a **structural reduction** in the supply of overnight unsecured money market funds: to this aim we decrease the value of all payment

[2] The BoF-PSS2 is a software developed by the Bank of Finland which allows the user to replicate (closely) realistic a settlement process and record a variety of statistics from it. It has been widely used to analyse the functioning of RTGS system, investigating eg how operational outages affect system performance. For a detailed introduction of the BoF-PSS2, see Leinonen and Soramaki (2003).

[3] From a theoretical point of view, an intraday money market could also compete with daylight central bank credit as a liquidity source in a RTGS environment, but on a practical grounds overnight is the shortest quoted maturity in the Euro money market. Therefore if intraday prices for liquidity seem to emerge, this is due to systematic differences in hourly prices of o/n deposits, as reported by Baglioni and Monticini (2008) on the basis of tick-by-tick data from the Italian interbank market e-MID.

transactions related to overnight deposits collected on each working day; ie both new contracts and repayments of past overnight deposits. In a second set of simulations, it is assumed that on each working day the banks are able to collect only a certain proportion of the value of overnight deposits collected, whereas the repayments of deposits obtained on previous days are kept unchanged. This latter simulation exercise, where each working day is treated as a separate observation, provides quantitative indications for the resiliency of the RTGS system under a **temporary money market shock** occurring before the operational day begins. In all the scenarios, the size of the shock is made to vary from 10 to 90 per cent of the total value of trading in unsecured deposits uniformly across all participants.

The last two sets of simulations estimate the impact of analogous reductions (ie varying from 10 to 90 per cent) in the availability of the other funding sources: cash balances held by participants in their central bank accounts and daylight collateralized credit lines granted by the central banks. Besides their own interest, these simulations represent a benchmark against which one can evaluate the indicators recorded for the money market based scenarios.

The whole exercise is carried out with reference to two reserve maintenance periods: 12 November to 9 December 2008, during the acute phase of the crisis, and 11 November to 7 December 2009.

The paper contributes to the recent literature on the RTGS system based on simulation studies, investigating for the first time how trading activity in the overnight unsecured money market affects the functioning of RTGS systems. So far, analyses of the resiliency of RTGS systems to liquidity shocks have been carried out by simulating either uniform reductions in liquid balances held by the RTGS participants in their central bank accounts (Koponen and Soramaki, 1998), or operational outages affecting one or more major system participants (eg Bedford et al, 2005, Arnold et al, 2006, Hellqvist and Koskinen, 2005, Lublóy and Tanai, 2007, Glaser and Haene, 2007, Heijimans, 2007).

The paper is organized as follows. In section 7.2, the methodology developed to identify the overnight unsecured deposits from transactions settled in TARGET2-Banca d'Italia is described; Section 7.3 reports some figures on TARGET2 activity during the two reference periods. In Section 7.4 the virtual clone of TARGET2-Banca d'Italia is presented, spotting its differences via the real TARGET2 system. Section 7.5 describes the data and the scenarios. Section 7.6 presents empirical results, and Section 7.7 offers concluding remarks.

7.2 Identification of unsecured overnight deposits settled via TARGET2-Banca d'Italia

The Italian and a few other European banks typically conduct overnight unsecured transactions either over the counter or through the e-MID electronic platform.[4] Both types of transactions are routinely settled via TARGET2. While pieces of information on e-MID transactions are readily available at Banca d'Italia, which is in charge of supervising the e-MID, OTC transactions can only be inferred from TARGET2 data.

Therefore, also relying on e-MID data to validate the results, we identify overnight unsecured transactions directly from TARGET2 settlement data, applying an identification methodology originally introduced by Furfine (1999) with reference to overnight deals settled via the US settlement system Fedwire.

The Furfine identification strategy basically compares the payment transactions sent by one participant to another on a given operational day, say T, with those sent by the latter to the former on the following day, T+1: if the difference between the payment settled at T+1 and that settled at T corresponds to a reasonable overnight interest amount, the payment pair is considered as stemming from a money market deal with maturity one-day (at either overnight, tom-next, or spot-next maturity) initiated in T and repaid in T+1.[5]

As pointed out by Demiralp et al (2004), with reference to data from the US RTGS system Fedwire, such an identification strategy embodies a trade-off between Type I and Type II errors: more specifically, under the null hypothesis that a Fedwire transfer is not an overnight loan, false positive identifications are Type I errors and false rejections of genuine loans Type II errors.

To minimize this potential source of misclassification, Furfine applies additional filters requiring that payment pairs which are loan

[4] The e-MID is the Italian screen based uncollateralized money market, which is organized and managed by e-MID SpA, a private company currently owned by banks and financial institutions. After years of continued growth, from the outset of the financial crisis, its market turnover has shrunk significantly due also to the banks' preference for over-the-counter trades (Bank of Italy, 2010).

[5] It is not possible to separate overnight contracts from those with the same maturity but with a postponed value date.

‘candidates’ exceed a minimum threshold and involve a round lot increment.[6]

Accordingly, in this paper we apply the Furfine algorithm to transactions settled in TARGET2-Banca d’Italia, filtering out all payments exceeding a threshold of €1 million and involving a round lot increment of €10,000. We consider as reasonable interest an overnight interest rate which falls within the range of the ECB standing facilities corridor, enlarged by 50 basis point (25 b.ps. on each side of the corridor).[7]

This methodology results in detection of more than 6,800 contracts amounting at €225 billion during the reserve maintenance period from 12 November to 9 December 2008, and more than 4,900 contracts amounting at €193 billion during the reference period from 11 November to 7 December 2009.

The availability of data on overnight deals traded on the e-MID platform[8] allows us to check the robustness of the results of our identification strategy, by mapping the e-MID contracts into the TARGET2 payment pairs marked as overnight deals by the algorithm.

A first check shows that we were able to correctly identify virtually all the e-MID contracts exchanged between Italian banks and around the 93 per cent of the e-MID deals traded by Italian banks and foreign counterparts.

A second check was carried out by comparing the densities of the reported e-MID overnight rates with those of the implied rates calculated over the TARGET2 transaction pairs classified as ‘overnight contracts’ by our identification procedure and as ‘OTC’ after the mapping exercise carried out as a first check. Chart 7.1 shows a substantial overlap between the two densities for both the reference periods, although during the first period the distribution of OTC rates is shifted leftward.[9]

---

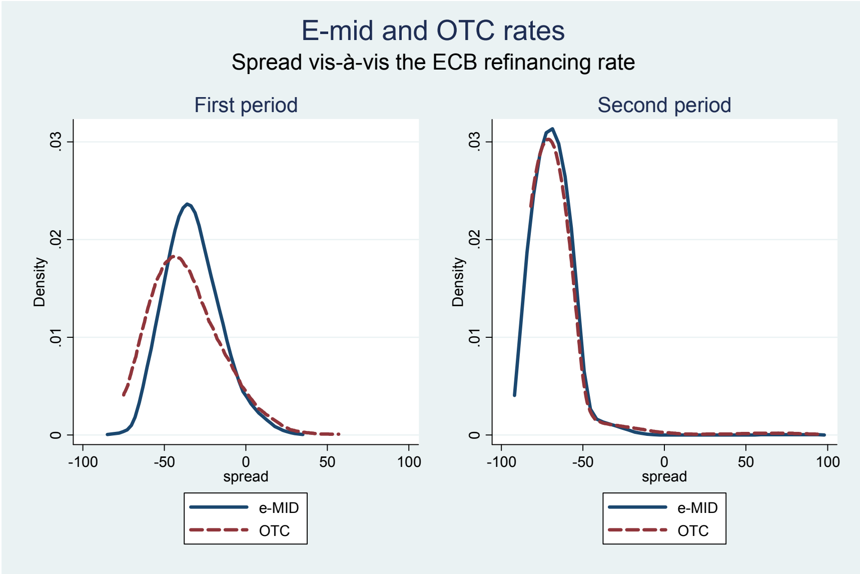
[6] Demiralp et al (2004) also evaluate whether the interest rate could have plausibly been a quoted rate in the market.

[7] The enlargement of the lower side of the corridor represents a robustness check. Because it is rare that banks are willing to trade deposits at a price less than the deposit facility, the extremely low number of identified contracts whose price falls under the deposit facility rate represents a positive result.

[8] The e-MID data are collected by the Bank of Italy for supervisory reasons and include all transactions executed on the platform.

[9] This difference might depend on the specific features of either the contracts (average size) or the intermediaries. It is not to be seen in any causal sense.

Chart 7.1 **Kernel densities**



The last check compares – for a period exceeding the two reference periods – the value of overnight deals reported by Italian banks included in the EONIA[10] panel and the value of overnight contracts as identified by our Furfine-like procedure. The comparison – not shown for confidentiality reasons – highlights a substantial overlap between the two variables, with a positive correlation of 80%.[11]

[10] The EONIA (Euro OverNight Index Average) is an effective overnight rate computed as a weighted average of all overnight unsecured lending transactions in the interbank market, initiated within the Euro area by a panel of contributing banks.

[11] In interpreting these figures, one should keep in mind that even with perfect identification the two variables are likely to differ, as in the EONIA panel only the overnight unsecured contracts negotiated by the panel banks on their account are reported while other contracts with one day maturity (ie spot and tom-next) and exchanges settled in TARGET2 on behalf of other banks are excluded.

7.3 The institutional context and the TARGET2-Banca d'Italia activity during the reference periods

TARGET2 represents the second generation of Trans-European Automated Real-time Gross settlement Express Transfer systems (TARGET) designed by the Eurosystem at the outset of the monetary union to facilitate integration of the money market in euro in order to allow for the smooth implementation of the single monetary policy and to improve the soundness and efficiency of the settlement of wholesale payments across national borders.

Whereas the first generation of TARGET had a decentralized technical structure consisting of the national RTGS systems and the ECB payment mechanism (EPM), interlinked to provide a technical infrastructure for processing euro denominated payments across national borders, TARGET2 – which started operations on 19 November 2007 – relies on a single technical platform, the 'Single Shared Platform' (SSP), jointly operated by three Eurosystem central banks, Banca d'Italia, Banque de France, and Deutsche Bundesbank.

To settle their payment obligations, TARGET2 participants can use their minimum reserve holdings during the day and access the unlimited, but collateralized, daylight credit facilities granted by the Eurosystem. In addition, the smooth functioning of the system is facilitated by a set of advanced functionalities (bilateral and multilateral debit limits, liquidity pooling, liquidity reservations, optimizations algorithms) enabling banks to manage effectively their payment flows throughout the business day.

Despite its technically integrated nature, from a legal point of view TARGET2 is a 'decentralized multiple system', ie which each participating and connected central bank is responsible for the operation of its own system component, for designing the system for finality, for maintaining business relationships with domestic participants, and for overseeing the local features of the system.

The simulation exercise was carried out in respect of the Italian component TARGET2-Banca d'Italia, which is one of the largest in the system, in terms of both volume and value of settlements.

A glance at TARGET2-Banca d'Italia operations over the two reference periods highlights the more favorable overall liquidity of system participants during the second period.

TARGET2 – Banca d'Italia participants settled average daily transactions amounting to €128 and €122 billion during the first and the second period, respectively. Despite the lower system turnover, in the second period the Italian participants held higher liquid balances on average during the whole operational day (67 versus €56 billion), partly as a result of an increase in collateral posted at the Bank of Italy; finally, on average, the share of unsecured overnight money market transactions was roughly the same in the two reference periods, amounting to 8 per cent of total interbank outgoing activity.

These figures show that the intermediaries have changed the way they managed intraday liquidity during the crisis, by modifying the mix of liquidity sources; more specifically, banks have relied more and more on Eurosystem refinancing operations and daylight credit lines and less on market liquidity.

Table 7.1 **TARGET2-Banca d'Italia (daily averages)**

	Maintenance period	
	12/11/2008–9/12/2008	11/11/2009–7/12/2009
Payment turnover (in EUR billion)		
Outgoing paymenta (a)	128.749	122.876
o.w.: cross border	43.564	45.710
o.w.: unsecured overnight money market transactions	11.300	10.025
Total liquidity		
Average balances (b)	56.564	66.966
Turnover (b/a) (%)	45	56
Daylight credit lines		
Collateral posted at the Bank of Italy	29.208	43.695
Average daylight credit usage	3.797	8.302

7.4 Simulator calibration

To carry out our simulation exercise, we relied on the *Bank of Finland Payment and Settlement System Simulator* (BoF-PSS2), developed by the Bank of Finland and currently used by some 80 institutions, eg central banks, universities, and clearing and settlement organizations, for oversight, research and operational purposes.

By combining algorithms and modules, BoF-PSS2 allows the user to create fictitious settlement systems, virtual clones of true systems, run counterfactual analyses, and record a variety of statistics from it.

In recent years, Bof-PSS2 has proven to be a powerful tool to investigate risk and efficiency issues of payment systems. Early studies estimate the trade-off between settlement delay and liquidity needs of different types of systems (Koponen and Soramäki, 1998, and Leinonen and Soramäki, 2005, among others); others quantify the systemic impact of operational outages at either participant or system level (Bedford et al, 2005, Ledrut, 2007, among others) or assess the efficiency of RTGS systems under different designs (Johnson et al, 2004).

Analyzing the effects of liquidity shocks on a payment system using the simulator requires *prima facie* replication of the functioning of a given payment system under normal conditions by calibrating the algorithms and parameters which define the system design in the simulator; the aim being to obtain the best possible fit between the pattern of payment flows in the resulting benchmark (or 'baseline') simulation[12] and the historical settlement data.

The version of the BoF-PSS2 used in this study is not able to perfectly replicate the design of the pan-European TARGET2 system due to the lack of some proprietary algorithms embodied in it. We therefore proceed to build a battery of potential virtual clones: among them, the best performing one in terms of ability to replicate a true business day of TARGET2-Banca d'Italia embeds an algorithm which releases queued payments on a *First Available – First Out* (FAFO) basis and an optimization mechanism which tries to settle all payments in the participants' queues on net basis every 20 minutes.

Our virtual clone exhibits very small deviations between its operational performances and those of the true TARGET2 Banca d'Italia system for the benchmark simulation:[13] i) all payments that were actually settled in TARGET2-Banca d'Italia during the reference period are settled using the artificial system; ii) a negligible number of transactions (a daily average of ten or 0.7 percent in value terms) experience a small increase in queuing time (on average, they remain in the queue for one more minute than that actually recorded in TARGET2-Banca d'Italia).

Other deviations from the real world are due to a lack of data on the level of reserves and collateral facilities available to TARGET2 participants holding settlement accounts at other Euro area national central banks. The chosen solution was to endow these foreign banks with unlimited liquidity, so as to insulate them from the shock effects. This solution implies that we do not take into account second round effects for the foreign banks: ie number and value of payments which banks located abroad will not be able to settle due to a lack of funds

[12] We define the 'benchmark simulation' as a simulation run with actual data drawn from TARGET2-Banca d'Italia.

[13] This does not necessarily imply that a plain vanilla RTGS system is as efficient as TARGET2, but only that for the level of liquidity at participants' disposal during the two reference periods, the liquidity saving mechanisms of TARGET2 are less influential.

caused by the fact that they did not receive the expected amount of overnight unsecured deposits from their Italian counterparts.

An analogous solution was adopted to deal with some ‘special’ participants, notably an Italian branch of a foreign bank which relies on the TARGET2 *liquidity pooling – virtual account facility*,[14] the Bank of Italy, and the Italian ancillary systems, Monte Titoli and Cassa Compensazione and Garanzia, ie the Italian central securities depositor and the central counterparty, respectively.

For ease of exposition, we postpone the submission time of the payments in the simulator until the actual settlement time historically recorded: this implies that every statistic on queued payments for every counterfactual scenario has to be read as the additional time that payments spend in a queue due to the simulated shock. Should the virtual clone be a perfect copy of TARGET2-Banca d’Italia, this would have implied that in the benchmark simulations no payments would have been placed in queue. However, due the (slight) deviations discussed above, even in the benchmark simulation, a few payments are not immediately settled but are funneled into participants’ queues.

7.5 Scenario description

We analyzed system-performance impacts of four types of shocks, ie structural and temporary reductions in overnight unsecured money market exchanges (hereafter, money market scenarios), and reductions in opening cash balances and in daylight credit lines, via five counterfactual scenarios that simulate the different shocks uniformly affecting the system participants. All the scenarios are evaluated against the benchmark simulation run with data historically recorded during the business days of the two reference periods.

In the structural reduction scenarios, it is assumed that on each working day banks are able to collect only a certain proportion of the value of the unsecured overnight deposits taken in. The size of the shock is made to vary from 10 to 90 per cent of the total value of trading in overnight unsecured deposits, reducing both the initial

[14] Under the liquidity pooling facility, a group of banks may aggregate on an intraday basis all the liquidity available in all their single TARGET2 accounts into one virtual account, which can be managed on a consolidated basis. Each transaction involving an account belonging to banks whose liquidity is pooled is immediately booked on the relevant single account using the global liquidity available in the virtual account.

transfers and the repayments. The other payments submitted to TARGET2 are kept unchanged. It is also assumed that the payments that remain unsettled at the end of the day are cancelled and not resubmitted the next business day.[15] On the first simulation day, the opening balances are those historically recorded, while on the following days the initial cash balances are kept equal to the previous end of day balances as calculated by the simulator.

In the alternative simulations based on the temporary shock hypothesis, a same size shock is considered but the repayments of the deposits obtained on the previous days are kept unchanged.

It is worth noting that the results of the two sets of simulation are not immediately comparable: in the former case, assessment of the shock impact must take into account the whole reference period; in the latter, each business day has to be considered a single observation.

Finally, the last two sets of simulations evaluate the impacts of same size uniform reductions in cash balances held by participants at the beginning of the business day and in the value of collateral posted at the Bank of Italy, respectively.

The quantification of the impact of each counterfactual scenario is based on a subset of the wide range of indicators which the BoF-PSS2 makes available to users as output simulation.[16]

The first set of indicators includes the volume and value (in both absolute and relative value) of payments which would have not been settled due to a liquidity shock. This set of indicators is of utmost importance since the inability of a bank to settle its payment obligations before the end of the value date is likely to entail a significant cost, either in term of explicit fees agreed with their customers or in terms of reputation.

The second set of indicators includes information on payment transactions that, due to the liquidity shock, would have been put into a queue and settled later on. They are

1. The *Maximum queue value*, ie the peak queue value during the business day
2. the *Average queue length*, ie the average queue duration of queued payments, namely the total queuing time of payments divided by the total number of queued transactions
3. the *Number of queued transactions*.

---

[15] It is implicitly assumed that the payments which remain unsettled at the close of the business day are redirected to alternative settlement arrangements.

[16] Bank of Finland (2009).

As a rule, payments entering a queue but settled before the close of the business day are less critical than payments that are cancelled by the system due to unavailability of funds. However, the common practice of having bilateral agreements among participants, industry guidelines and system provisions for the settlement of specific payment types (eg cut-offs for settlement of time critical payments stemming from ancillary systems) may give rise to explicit or implicit settlement delay costs. Indicators related to the value and volume of queued payments and the time they spend in queues may also be viewed as proxies of potential tensions in the money market, as they could generate a greater demand for funds to meet specific time critical cut-off times and may, in extreme circumstances, increase the interest rate volatility.

A third set of indicators is related to the participants' liquidity needs and comprises:

1. the *daylight credit usage* ie the percentage of daylight credit lines actually used by the participants to meet their payments obligation
2. the *Lower liquidity bound*, ie the minimum amount of liquidity required to settle all payments submitted during a day.[17]

7.6 Simulation outcomes

Unsettled payments. The output related to the unsettled payments shows a high degree of resiliency of the system, as the participants would have met virtually all their payment obligations before the close of the business day even in the event of an extreme structural reduction in the supply of overnight unsecured money market deposits.

In the first maintenance period considered, a contraction of 50 per cent in unsecured overnight loans would have increased the quantity of unsettled payments to not more than €350 million, equal to 1 per cent of the total. In the second period, the increase in the value of unsettled transactions would have amounted at 50 million for a same size shock. In all the money market scenarios the amounts not settled would have amounted to an extremely small proportion of those

[17] Ie the liquidity the banks need to hold is just enough to net-settle their payments before the end of the day by applying multilateral offsetting. For a more detailed description of the Lower liquidity bound indicator, see Bank of Finland (2009).

entered into the system[18] (see table A7.1a in the Appendix). A corollary of these results is that the contagion between TARGET2 and other settlement arrangements would likely have been limited: if TARGET2 Italian participants had redirected the payments not settled in the gross settlement to either the competing system EURO1 or to the correspondent bank, the impact on these arrangements would have been limited in terms of both liquidity and risk.

The temporary shock scenarios highlight a lower resiliency of the system: for a 50 per cent contraction, the unsettled payments would have exceeded €700 million, twice those recorded for the previous simulations. In the second period, the differences between the two sets of simulations would have been significantly smaller.

In all money market scenarios, given the extremely high number of low value payment channeled in the system, the volume not settled would have been extremely small relative to those entered into the system (from less than one payment per day to a maximum of 85 payments per day for reduction in the supply of unsecured overnight deposits, ranging from 10 to 90 per cent), but of course their *per capita* value would have been relatively high.

Same size reductions in the level of either the opening cash balances or the daylight credit lines would have generated different impacts on the unsettled payments. In both the reference periods, the effect of a contraction in daylight credit lines would have been nil,[19] whereas a shrinkage in the opening cash balances would cause a positive, albeit small, amount of unsettled transactions (table A7.1b).

Queued payments. The effects for queued payments would have been more substantial, even for relatively mild scenarios entailing the risk of potential congestion in the system (table A7.2a). Again, the effects of the shocks in the second maintenance period would have been smaller.

The scenarios based on the temporary shocks show more pronounced effects: more specifically, even in the event of a

[18] It is worth recalling that – as mentioned above – the non-Italian banks are endowed with unlimited liquidity in the simulations. This implies that our model does not fully capture some indirect effects related to the contagion from Italian to non-Italian banks joining other TARGET2 national components. All in all, the relevance of these second round effects does not appear extremely significant since TARGET2-Banca d'Italia represents less than 10 per cent of the total TARGET2 turnover.

[19] Since the daylight credit granted by the central bank is routinely repaid by banks before the end of the business day, its reduction unlikely gives rise to unsettled transactions. During the two reference periods, this event occurred only once on the very last day of the first maintenance period.

contraction of 10 per cent, the number, value , and delay time of queued payments would have increased significantly, especially during the first reference period (from 10 to 127 payments per day, for average queuing times of 1 and 10 minutes respectively). The peak value of queued payment would have followed a similar path.

In the first reference period, a contraction of 50 per cent in unsecured overnight loans would have led to more than 600 payment transactions funnelled into the queues every day for an average queuing time of more than one hour; at their peak, the queued payments would have increased by a daily average of more than €8 billion.

Much in the same way, a significant degree of congestion would have emerged in the event of mild shocks to cash balances and daylight credit lines. As expected, given the heavier recourse to the central bank credit, reductions of the value of collateral would have resulted in more pronounced effects during the second reference period (table. A7.2b).

Liquidity needs. The credit usage indicator appears to increase only slightly for money market shocks of increasing size, confirming that the majority of banks tend to pledge collateral at the central bank in excess respect their normal funding needs. Similar, the lower liquidity bounds indicator exhibits small increases, relatively higher in case of temporary shocks (table A7.3a).[20]

7.7 Conclusions

From a policy point of view, the results of the exercise, which need to be assessed in light of the above methodological simplifications in calibrating the simulator, show that even a drastic fall in the supply of overnight unsecured deposits would not have dramatically impaired the functioning of the system, as the other funding sources available to the participants would have allowed the banks to settle virtually all of their payment obligations before the end of the business day, without any additional interventions by the Eurosystem.

[20] This last set of indicators is not computed for the cash balance scenario or the daylight credit line scenarios since a) the *lower liquidity bound* varies only in response to changes in the total daily amount of payments to be settled; b) the credit usage would be directly affected by the reduction on the value of posted collateral.

In both the reference periods, there might have been more substantial effects for queued payments, even in milder scenarios, which would have altered significantly the intraday pattern of payments. A significant amount of queued payments represents an important performance indicator for an RTGS system, since banks which do not receive payments when expected are likely to postpone submission of their payments into the system even if they have sufficient liquidity in their accounts, thus aggravating the congestion of the system; such a scenario might evolve into a gridlock[21] in extreme circumstances.

The effect of a money market shock would have been significantly smaller in the second reference period due to both a lower ratio of unsecured overnight loans to the total value of settled payments and a higher average level of liquidity in participants' settlement accounts. This evidence suggests that a reduction in the supply of the unsecured overnight funds could have had more severe effects before the financial turmoil, when participants' relied to a great extent on the unsecured money market.

When we come to the objective of gaining a better understanding of the role played by the various liquidity sources, it emerges that the impact of temporary shocks in the unsecured overnight market would have been of similar magnitude to those stemming from a reduction in opening cash balances of participants at the Banca d'Italia. Despite the fact that the money market ensures only a redistribution of existent liquidity, it may play a role analogous to those of the other liquidity funds.

Moreover, with specific reference to the money market, we observe that structural reductions would have impacted the functioning of the RTGS system less than the temporary shocks, a result that is fully consistent with trading activity in the unsecured money market characterized by a high degree of rollover, where a significant share of the trading in the day T is carried out to fund the repayment of the deposits traded on the previous day.

This paper may be extended in several ways: first, the exercise could be run with the inclusion of 'pre-crisis' periods to compare outcomes for different money market conditions; a further extension might be to drop the hypothesis of uniform reductions by allocating randomly the x-per cent shrinkages to the participants.

[21] A gridlock is 'a situation that can arise in a funds or securities transfer system in which the failure of some transfer instructions to be executed (because the necessary funds or securities balances are unavailable) prevents a substantial number of other instructions from other participants from being executed' BIS (2003).

Appendix

Table A7.1a **Unsettled payments. Shocks on the overnight unsecured money market (daily averages)**

Contraction size (%)	Structural reduction: Value of unsettled payments (eur million)	Structural reduction: Value of unsettled payment (% of total interbank payments)	Structural reduction: Volume of unsettled payments (in units)	Temporary shock: Value of unsettled payments (eur million)	Temporary shock: Value of unsettled payment (% of total interbank payments)	Temporary shock: Volume of unsettled payments (in units)
First round						
Baseline	0	0.00	0	0	0.00	0
10	5	0.00	0	36	0.04	1
30	124	0.13	4	340	0.34	6
50	357	0.38	13	785	0.81	15
70	666	0.74	22	1,436	1.51	34
90	965	1.10	32	2,069	2.22	85
Second round						
Baseline	0	0.00	0	0	0.00	0
10	20	0.02	0	35	0.03	1
30	26	0.03	1	61	0.07	3
50	45	0.05	1	90	0.10	7
70	73	0.08	3	139	0.15	9
90	101	0.12	3	188	0.22	17

Table A7.1b **Unsettled payments. Cash balances and daylight credit lines scenarios (daily averages)**

	Cash balances			Daylight credit lines		
Contraction size (%)	Value of unsettled payments (eur million)	Value of unsettled payment (% of total interbank payments)	Volume of unsettled payments (in units)	Value of unsettled payments (eur million)	Value of unsettled payment (% of total interbank payments)	Volume of unsettled payments (in units)
			First round			
Baseline	0	0.00	0	0	0.00	0
10	388	0.38	1	14	0.00	0
30	661	0.65	4	14	0.00	0
50	1,211	1.19	7	14	0.00	0
70	1,867	1.84	27	14	0.00	0
90	2,746	2.71	71	14	0.00	0
			Second round			
Baseline	0	0.00	0	0	0.00	0
10	393	0.42	1	0	0.00	0
30	513	0.55	3	0	0.00	0
50	806	0.87	9	0	0.00	0
70	1,048	1.13	20	0	0.00	0
90	1,676	1.81	32	0	0.00	0

Table A7.2a **Queued payments. Shocks on the overnight unsecured money market (daily averages)**

Contraction size (%)	Structural reduction: Maximum queue value (eur million)	Structural reduction: Average queue length (minute)	Structural reduction: Number of queued payments (in units)	Temporary shock: Maximum queue value (eur million)	Temporary shock: Average queue length (minute)	Temporary shock: Number of queued payments (in units)
			First round			
Baseline	0	0	10	10	0	10
10	201	11	45	589	12	127
30	1,121	39	180	3,838	15	361
50	3,055	46	310	8,521	25	622
70	5,179	41	493	15,444	27	767
90	7,292	46	794	22,940	37	919
			Second round			
Baseline	2,200	32	3	2,200	32	3
10	1,936	18	17	2,542	29	13
30	2,138	17	43	2,820	30	40
50	2,246	15	52	3,151	37	56
70	2,795	28	86	3,615	29	88
90	2,980	20	122	4,023	40	127

Table A7.2b **Queued payments. Cash balances and daylight credit lines scenarios (daily averages)**

Contraction size (%)	Cash balances: Maximum queue value (eur million)	Cash balances: Average queue length (minute)	Cash balances: Number of queued payments (in units)	Daylight credit lines: Maximum queue value (eur million)	Daylight credit lines: Average queue length (minute)	Daylight credit lines: Number of queued payments (in units)
			First round			
Baseline	0	0	10	10	0	10
10	9,544	26	31	4,198	10	77
30	11,013	16	160	4,818	8	312
50	18,620	13	420	6,473	10	567
70	24,963	24	1,053	9,035	8	930
90	34,944	28	1,909	10,721	7	1,984
			Second round			
Baseline	2,200	32	3	2,200	32	3
10	11,109	27	27	7,348	10	44
30	13,642	18	86	8,883	4	418
50	16,918	12	318	10,615	4	771
70	20,907	13	633	14,793	3	1,220
90	27,092	10	1,297	16,786	5	1,915

Table A7.3a **Liquidity needs. Shocks on the overnight unsecured money market (daily averages)**

	Structural reduction		Temporary shock	
Contraction size (%)	Daylight credit usage* (% actually used)	Lower Liquidity Bound (eur million)	Daylight credit usage* (% actually used)	Lower Liquidity Bound (eur million)
		First period		
Baseline	0.0	4,417	0.0	4,417
10	0.0	4,437	0.6	4,525
30	0.1	4,611	1.0	4,852
50	0.5	4,849	1.6	5,165
70	1.1	5,108	2.3	5,574
90	1.8	5,306	2.9	6,008
		Second period		
Baseline	0.0	4,821	0.0	4,821
10	0.2	4,825	0.0	5,186
30	0.2	4,921	0.3	5,457
50	0.2	4,923	0.7	6,142
70	0.3	5,047	1.2	6,940
90	0.5	6,249	1.9	7,808

* Changes respect the baseline scenario.

References

Angelini, P (1998) **An analysis of competitive externalities in gross settlement systems.** Journal of Banking & Finance, Elsevier, Vol. 22(1), 1–18, January.

Arnold, J – Bech, M L – Beyeler, W E – Glass, R J – Soramäki, K (2006) **The topology of interbank payment flows.** Federal Reserve Bank of New York, Staff Report 243.

Ashcraft, A B – Duffie, D (2007) **Systemic Illiquidity in the Federal Funds Market.** American Economic Review, American Economic Association, Vol. 97(2), 221–225, May.

Arciero, L – Biancotti, C – D'Aurizio, L – Impenna, C (2008) **Exploring agent-based methods for the analysis of payment systems: a crisis model for StarLogo TNG.** Temi di discussione (Economic working papers) 686, Bank of Italy, Economic Research Department

Bank of Finland (2009) **BoF-PSS2 User Manual version 2.4.0.** August.

Baglioni, A – Monticini, A (2008) **The intraday price of money: evidence from the e-MID market.** Journal of Money, Credit and Banking, Blackwell Publishing, Vol. 40(7), 1533–1540, October.

Bank for International Settlement (1997) **Report on real-time gross settlement systems.** Basle.

Bank for International Settlement (2003) **A glossary of terms used in payments and settlement systems.** Basle.

Bank of Italy (2010) **Financial Stability Report No. 1.** December.

Bech, M – Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory, 109: 198–219.

Bedford, P – Millard, S – Yang, J (2005) **Analysing the impact of operational incidents in large value payment systems: a simulation approach.** In Leinonen, H (ed.), Liquidity, risks and speed in payments and settlement systems – a simulation approach, Bank of Finland Studies, E:31.

Demiralp, S – Preslopsky, B – Whitesell, W (2006) **Overnight interbank loan markets.** Journal of Economics and Business, Elsevier, Vol. 58(1), 67–83.

Furfine, C H (1999) **Interbank exposures: quantifying the risk of contagion.** BIS Working Papers 70, Bank for International Settlements.

Glaser, M – Haene, P (2007) **Simulation of participant-level operational disruption in Swiss Interbank Clearing.** presentation given at the 5th simulator workshop in Helsinki, http://www.bof.fi/en/rahoitusmarkkinat/kehityshankkeet/BoF-PSS2/seminar07.htm.

Heijimans, R (2007) **Stress simulations: a Dutch case.** Presentation given at the 5th simulator workshop in Helsinki, http://www.bof.fi/en/rahoitusmarkkinat/kehityshankkeet/ BoFPSS2/ seminar07.htm.

Hellqvist, M – Koskinen, J (2005) **Stress testing securities clearing and settlement systems using simulations.** In Leinonen, H (ed.), Liquidity, risks and speed in payments and settlement systems – a simulation approach, Bank of Finland Studies, E:31.

Klee, E (2010) **Operational outages and aggregate uncertainty in the federal funds market.** Journal of Banking & Finance 34, March.

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

Johnson, K – McAndrews, J – Soramäki, K (2004) **Economizing on Liquidity with Deferred Settlement Mechanisms.** FRBNY Economic Policy Review, December.

Ledrut, E (2007) **Simulating retaliation in payment systems: Can banks control their exposure to a failing participant?** DNB Working Papers No. 133, April.

Leinonen, H – Soramäki, K (2003) **Simulating interbank payment and securities settlement mechanisms with the BoF-PSS2 simulator.** Bank of Finland Discussion Paper 23/2003, Bank of Finland. Research Department.

Lublóy, A – Tanai, E (2007) **Operational disruption and the Hungarian RTGS system VIBER.** Presentation given at the 5th simulator workshop in Helsinki, http://www.bof.fi/en/rahoitusmarkkinat/kehityshankkeet/BoF-PSS2/seminar07.htm.

Chapter 8

Participant operational disruptions: the impact of system design

Ashwin Clarke – Jennifer Hancock**

* We are grateful to Mark Manning, Andrea Brischetto, Christopher Kent, other colleagues at the Reserve Bank of Australia, and participants at Bank of Finland's 2009 Payment and Settlement System Simulation Seminar for their helpful comments. We would also like to thank Sammy Yousef, Rod Phillips and the Bank of Finland staff for their technical advice. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Reserve Bank of Australia. The authors are solely responsible for any errors.
Authors: clarkea or hancockj at domain rba.gov.au
Media Office: rbainfo@rba.gov.au

8 Participant operational disruptions: the impact of system design

Abstract

Real-time gross settlement (RTGS) systems often incorporate elements of net settlement systems (‘hybrid features’) to economise on liquidity. Such hybrid features could mitigate the systemic impact of participant operational disruptions. However, participants may decrease their holdings of liquidity in response to the inclusion of such hybrid features, thus potentially negating the benefit of such features in the event of an operational disruption. This paper simulates participant operational disruptions, using data from Australia’s RTGS system, the Reserve Bank Information and Transfer System (RITS), to analyse the effect of system design on the systemic impact of such disruptions. Five different system designs are analysed in an attempt to isolate the effect of including a central queue, a liquidity saving algorithm and a liquidity reservation feature. The results suggest that the liquidity saving algorithm and liquidity reservation features in RITS generally mitigate the impact of a participant’s operational disruption. The liquidity reservation feature in RITS reduces the systemic impact of an operational disruption as long as participants react to the disruption by stopping payments to the stricken participant and using reserved liquidity. The liquidity saving algorithm also mitigates the impact of a disruption, even if participants do not react. In combination, these hybrid features mitigate the systemic impact of a disruption, even if participants choose to decrease the liquidity they commit to the RTGS system when the system design incorporates a liquidity saving algorithm. The hybrid features also tend to reduce the sensitivity of the system to the size of the participant with the operational disruption.

8.1 Introduction

High-value payment systems are critical infrastructure for financial markets. To mitigate the systemic impact of a participant’s default, most high-value payment systems now settle on an RTGS basis (Bech, Preisig and Soramäki, 2008). But while RTGS eliminates credit risk between participants, it is liquidity intensive since payments are settled individually. To limit the call on participants' collateral to secure additional intraday liquidity it is important that liquidity recycles efficiently through the system. If an operational disruption results in that participant being unable to send payment instructions to the RTGS system for settlement, liquidity from its receipts accumulates in the stricken participant's account, forming what is known as a 'liquidity sink'. Such a disruption in liquidity recycling can prevent other participants from settling their payments.

The design of RTGS systems vary significantly around the world. Many RTGS systems incorporate elements of net settlement systems to economise on liquidity. Such hybrid features could mitigate the systemic effect of participants' operational disruptions. Glaser and Haene (2008) suggest that a central queue can reduce the size of the liquidity sink that results from a participant's operational disruption because the payments already queued by the stricken participant can still settle. Also, since a liquidity saving algorithm minimises the amount of liquidity required to settle payments, these mechanisms can potentially reduce the value of unsettled payments that result from any liquidity shortage caused by a participant's operational disruption. Furthermore, since liquidity reservation features tend to slow liquidity recycling, they can potentially minimise the systemic impact of a participant's operational disruption by slowing the flow of liquidity into the stricken participant's account.

Working in the opposite direction, however, unless the liquidity reservation feature specifically targets the stricken participant (ie there are bilateral limits) it can slow payments between all participants. By restricting the flow of liquidity, reserving liquidity may increase the value of unsettled payments, including payments not involving the stricken participant. In addition, the presence of a liquidity saving algorithm may result in participants decreasing the liquidity they commit to the RTGS system, thus negating the benefit these mechanisms might have during an operational disruption.

This paper analyses the effect of system design on the systemic impact of participant operational disruptions using a simulator developed by the Bank of Finland ('the simulator'). These simulations use data from Australia's RTGS system, RITS. As RITS features a central queue with a bilateral offset algorithm, as well as a liquidity reservation feature, it provides a rich dataset with which to analyse the effects of system design. The paper also investigates how hybrid features interact with participant reaction times, and how they may alter the relationship between the size of the stricken participant and the systemic impact of an operational disruption.

As with all simulation studies, the lack of an endogenous behavioural response means that the results should be interpreted with care. In particular, simplifying assumptions are made regarding changes in participant behaviour in response to a change in system design.

The remainder of the paper is structured as follows. Section 8.2 provides an overview of the literature on system design and operational outages. Section 8.3 describes RITS and its liquidity saving and liquidity reservation features. Section 8.4 presents the methodology used to analyse the effect of system design on operational disruptions in RITS. Section 8.5 presents the results of the simulation and Section 8.6 concludes.

8.2 Literature review

In recent years there has been a sharp increase in payments settled in hybrid RTGS systems. In 1999, three per cent of the total value settled in large value payments systems was settled in RTGS systems that incorporated hybrid features; by 2005 this share had grown to roughly 32 per cent (Bech et al, 2008).[1] At the same time, hybrid systems have received increased attention in payments literature, for example: McAndrews and Trundle (2001) and BIS (2005) provide detailed expositions of hybrid systems; Johnson, McAndrews and Soramäki (2005) and Ercevik and Jackson (2009) use simulation analysis to quantify the impact of introducing hybrid features on liquidity demand and settlement delays; while Martin and McAndrews (2008), and Galbiati and Soramäki (2010) use theoretical models to analyse the impact on participants' incentives.

A separate stream of payments literature has focussed on analysing operational risk in RTGS systems through simulation studies. This literature generally follows the methodology established by Bedford, Millard and Yang (2005); to analyse the systemic effects of simulated operational disruptions that prevent a participant (or multiple participants) from submitting payments. Bedford et al simulate operational disruptions in the UK RTGS system, CHAPS, while Schmitz and Puhr (2007), Glaser and Haene (2008), Anderson and

[1] This is based on a study covering CPSS member countries (as at 2005) and non-CPSS euro area countries.

Madson (2009) and Lubloy and Tanai (2009) perform similar analyses on Austrian (ARTIS), Swiss (SIC), Danish (KRONOS) and Hungarian (VIBER) RTGS systems, respectively. Results from the simulation analysis vary between studies, with differences largely explained by the level of liquidity in the system under consideration and the size of the participant experiencing the disruption, as well as assumptions about unstricken participants' reaction to the disruption.

Ledrut (2007) investigates the mitigating effect of the participant reactions by simulating counterparties' reactions to a participant disruption in the Dutch RTGS system, TOP. Participants are assumed to react by stopping payments to the stricken participant after a pre-determined time has elapsed or once their exposure to the stricken participant has reached a certain threshold. Ledrut concludes that more timely participant reactions can significantly reduce the systemic consequences of participant level operational disruptions. Merrouche and Schanz (2009) also investigate counterparties' reactions to a participant operational disruption. Based on an econometric model of CHAPS, they find that payment flows to stricken participants tend to decrease until around one hour into the disruption, but increase slightly afterwards, presumably as the cost of violating contract obligations or market practices by delaying payment increases.

8.3 Australia's RTGS system

RITS has operated as an RTGS system since 1998.[2] Over 90 per cent of interbank settlements, by value, in Australia are settled on a gross basis through RITS;[3] this share has been broadly steady since RITS commenced operations in 1998. In 2008, RITS settled on average around 32,000 transactions each day, with a total average value of $186 billion, using around $17 billion of liquidity. Liquidity in RITS is sourced from overnight balances held in participants' accounts at the Reserve Bank of Australia (RBA) and additional funds made available to participants by the RBA via interest-free intraday repurchase agreements (repos). Access to these funds is limited only by participants' holdings of eligible securities. In the sample period, RITS had 59 direct participants, although the system is quite

[2] For more information on RTGS in Australia see Gallagher P, Gauntlett, J and Sunner, D (2010).

[3] The remaining 10 per cent of interbank settlements in RITS are settled in deferred net batches.

concentrated with the four major Australian banks counterparts to almost 60 per cent of transactions settled through RITS.

The central queue in RITS operates on a 'bypass first-in first-out (FIFO) basis'.[4] If the transaction being tested for settlement cannot be settled individually, the bilateral offset algorithm searches for up to 10 offsetting transactions (additively, in FIFO order), which it attempts to settle simultaneously.[5] RITS incorporates a liquidity reservation feature that allows participants to manage their payments and reserve liquidity for 'priority' payments. To facilitate this process, RITS participants have access to real-time information, including their settled and queued payments and receipts. The liquidity reservation feature in RITS allows participants to set a 'sub-limit', with balances below this limit reserved for settlement of priority transactions. In contrast, 'active' payments are only tested for settlement against balances in excess of the sub-limit, while 'deferred' payments are not tested for settlement until the sending participant changes the status of the payment to either active or priority. This can be done at any time prior to settlement.

Approximately 30 per cent of the value of RITS payments settled between January and April 2008 were settled using bilateral offset, while 25 per cent were settled as priority payments, using liquidity protected by sub-limits (Figure 8.1). The bulk of the value of payments in RITS is settled between 3.00pm and 5.00pm, during which time the majority of priority payments are made. In contrast, RITS volumes are concentrated at the beginning of the day, with a large number of small payments settling around 9.15am, immediately after the opening of the system.[6]

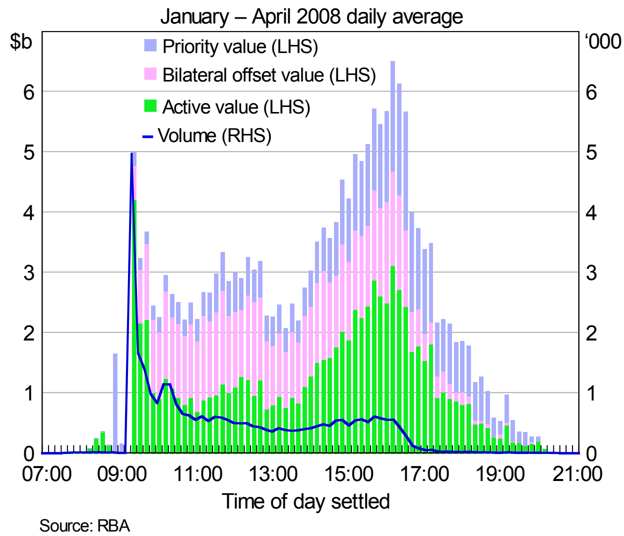
---

[4] Payments are tested for settlement in FIFO order, but rather than stopping if the first payment cannot be settled immediately the system moves on to test the next payment in the queue for settlement, and so on, looping back to the first payment when it reaches the end of the queue.

[5] In July 2009, the RBA added a Targeted Bilateral Offset algorithm, which allows participants to select specific payments for bilateral offset.

[6] Only payments associated with the settlement of the retail payments systems can be settled in RITS prior to 9.15am.

Figure 8.1 **Use of hybrid features in RITS**



8.4 Methodology

8.4.1 Data

The simulations are based on RITS transaction, liquidity and sub-limit data from 10 business days in the first quarter of 2008. This period represents a typical fortnight in 2008, during which an average of around 31,000 transactions worth $191 billion were settled each day using around $15 billion of liquidity.

8.4.2 System design

To measure the marginal benefit of hybrid features during an operational disruption, participant level disruptions are simulated under five different system designs (Table 8.1). Since hybrid features usually require a central queue, all system designs, other than the pure RTGS system, incorporate this feature. To roughly disentangle the effects of sub-limits and bilateral offset, a scenario with only sub limits and a scenario with only bilateral offset are examined. Rather than using the existing algorithms in the simulator, this paper uses modified algorithms that better replicate the bilateral offset and liquidity reservation features in RITS (See Appendix for further details).

Table 8.1 **System designs**

	Central queue	Bilateral offset	Sub-limits
Pure RTGS	–	–	–
RTGS with central queue only	x	–	–
RTGS with bilateral offset	x	x	–
RTGS with sub-limits	x	–	x
RITS replica	x	x	x

8.4.2.1 Submission times

While a change in system design is likely to provide an incentive for participants to vary submission times, for simplicity this paper generally assumes that there is no change in submission behaviour.[7] However, since there is no central queue to co ordinate payments in a pure RTGS system, participants require some internal mechanism to ensure that a payment is only sent to the RTGS system when the participant has sufficient funds to settle that payment. Consequently, RITS submission times are unlikely to be appropriate when simulating a pure RTGS system. Instead, settlement times from the benchmark simulations of the central queue-only system are used to proxy the submission times in the pure RTGS system.[8] As a result, the key difference between the pure RTGS and central queue-only simulations is the payments on the queue. Even so, this is likely to underestimate the benefits of a central queue since the visibility of queued receipts on a central queue can decrease participants' uncertainty regarding their future liquidity requirements and thereby reduce their incentive to delay submitting payments.[9]

[7] The submission times and status of payments whose status changes between submission and settlement in RITS have been amended to best replicate when these payments settle, as the option to change payment status is not available in the simulator. See Appendix for further details.

[8] In addition, as simulations cannot incorporate the re-submission of payments that do not settle immediately in a pure RTGS system, the central queue model (with adjusted submission times) is used to simulate the pure RTGS system. Payments that do not settle immediately are queued and re-tested for settlement at a later stage, as if they had been re-submitted.

[9] RITS provides each participant with real-time information on their queued payments and receipts. In order to prevent participants incurring credit risk by crediting their customers before interbank settlement has occurred, while the paying and receiving

The submission time assumptions are also likely to understate the benefit of a bilateral offset algorithm and sub-limit functionality, since inclusion of each of these features provides an incentive to submit payments earlier. A bilateral offset algorithm reduces the incentive to submit payments late by potentially minimising the amount of liquidity required to settle payments, especially in combination with sub-limits that allow participants to reserve liquidity for time critical payments.

8.4.2.2 Liquidity

As noted previously, however, participants may decrease their holdings of liquidity in response to the inclusion of liquidity saving algorithms, thus potentially negating the benefit of such features in the event of an operational disruption. Consequently, we report results both using actual liquidity from RITS and assuming that participants decrease their holdings of liquidity by 30 per cent in response to the presence of the liquidity saving algorithm.[10] In the latter simulations, participants' actual shares of liquidity have been maintained, as this should be a reasonable indicator of each participant's relative access to liquidity.[11]

A 30 per cent reduction in liquidity was selected after analysing the effect of varying available liquidity on the value of unsettled payments in each of the four system designs with a central queue (Figure 8.2). As expected, the bilateral offset algorithm significantly decreases the liquidity required to settle payments. However, an extremely large increase in liquidity would be required to settle all payments in those systems without the bilateral offset algorithm. Consequently, instead of scaling liquidity up in these systems, liquidity in the RITS replica and bilateral offset systems is scaled down by 30 per cent to equalise

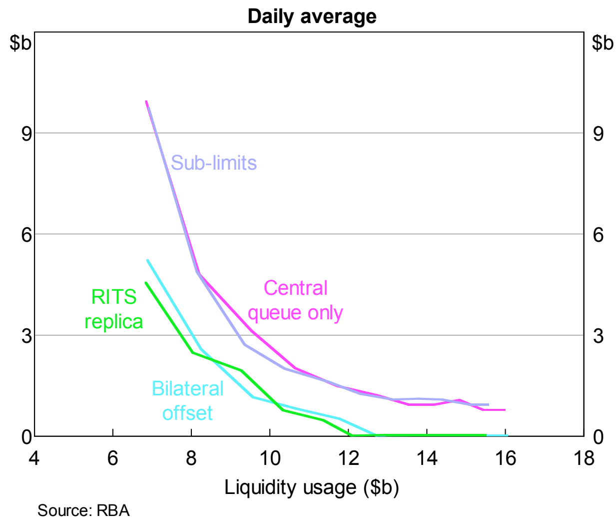
participant are identified the receiving participant does not receive details of the ultimate beneficiary until the payment has settled.

[10] In the 'actual liquidity' simulations we assume that participants do not unwind intraday repos until the end of the day to minimise the effects of the changes in the timing of settlement in the simulations. This is a reasonable assumption if the main driver of the cost of liquidity is the maximum value of collateral used, rather than the length of time during the day that the securities are used.

[11] The RBA, CLS Bank, and the settlement accounts of the equity and futures clearing and settlement systems are provided with unlimited funds.

the value of unsettled payments across all system designs at around \$1.5 billion.

Figure 8.2 **Unsettled payments**



8.4.2.3 Sub-limits

In general, observed behaviour (ie sub-limits and payment status) in RITS is replicated when simulating systems with sub-limit functionality. However, it is reasonable to assume that participants will use all available liquidity to settle any payments outstanding at the end of the day; therefore sub-limits are reduced to zero shortly before the system closes to allow as many payments to settle as possible.

8.4.3 Measuring the effect of a disruption

The primary statistic used to measure the impact of a disruption is the total value of unsettled payments.[12] Given that the focus is the systemic impact of a disruption, the measure of unsettled payments

[12] The impact of an operational disruption could have been measured in an equivalent fashion in terms of the value of additional liquidity required to settle all transactions. Another measure of the systemic impact is the simulator's settlement delay indicator.

excludes payments to or from the stricken participant.[13] In addition, the value of the liquidity sink, measured as the stricken participant's end-of-day balance after repaying any intraday repos, is reported.

8.4.4 Simulation scenarios

In analysing the interaction between system design and participant reaction time, this paper follows the methodology used by Glaser and Haene (2008), who build on the approach established by Bedford et al (2005), to find the time when the largest 'theoretical liquidity sink' will form in RITS.[14] The largest theoretical liquidity sink can be found by maximising the following equation:

$$\begin{aligned}&\text{Theoretical Liquidity Sink}_{it} = \\ &\text{Balance}_{it} + \sum\nolimits_{t}^{t+R} \text{Receipts}_{it} - \text{Value on Queue}_{it}\end{aligned} \tag{8.1}$$

where i is the stricken participant, Balance_{it} is participant i's balance at time t, Receipts_{it} is the value of its receipts, $\text{Value on Queue}_{it}$ is the value of outgoing payments on the queue and R represents the time it takes unstricken participants to react. When identifying the largest theoretical liquidity sink, it is assumed that non stricken participants take 2 hours to react and that the disruption to participant i's payments lasts until the end of the day.

An operational disruption at the participant and time identified using the method above is then simulated using each of the system designs, with participants reacting after 10 minutes, 2 hours or not reacting at all (Table 8.2). Since system design affects exactly when payments settle, the simulation starts from the point of the disruption to ensure that the results across systems are comparable. As a result, for any given day simulated, the same payments are outstanding at the start of the simulation, regardless of the system design being simulated.

[13] Note that changing the system design (without an endogenous response to this by participants) can result in unsettled payments due to insufficient liquidity, even without simulating a participant operational disruption. As a result, the value of unsettled payments, particularly in systems without bilateral offsetting, may be slightly overstated.
[14] In common with Bedford et al, the largest theoretical liquidity sink is restricted to the morning to ensure that there is a significant value of payments to settle after the disruption.

In systems with sub-limits, two reactions to the disruption are considered. In the first case, it is assumed that participants react by not only stopping payments to the stricken participant, but also dropping their sub-limits to zero to maximise the liquidity available to settle payments between unstricken participants.[15] In the second, it is assumed that participants do not drop their sub-limits.

Table 8.2 **Scenarios**

Scenario	Stricken participant	Time of disruption	Liquidity assumption	Reaction time	Sub-limit assumption	Number of scenarios
Benchmark	na	na	Actual	na	na	5
Analysis of reaction times	Largest theoretical liquidity sink		Actual Scaled	10 minutes 2 hours No reaction	Set to zero Unchanged	27
Analysis of participant size	Largest 15 participants	9.15am; 12.00pm; 3.00pm	Actual	2 hours	Set to zero	225

In addition to the scenarios above, this paper also investigates how system design affects the relationship between the size of the participant experiencing the operational disruption and the systemic effects of that disruption. This involves conducting a further set of simulations for the largest 15 participants (measured by value of payments submitted and received).[16] These simulations use a 2 hour reaction time, as anecdotal evidence suggests that this is the approximate time it takes participants in RITS to react to an operational disruption. As the value of queued payments varies at different times of day, and hence the impact of system design will vary, disruptions at 9.15am, 12.00pm and 3.00pm are modelled in these participant size scenarios.

[15] However, the simulator is unable to prevent priority payments submitted before the time at which participants react to the operational disruption from settling after participants react.

[16] Excluding the Reserve Bank, CLS Bank and the settlement accounts for the equity and futures markets.

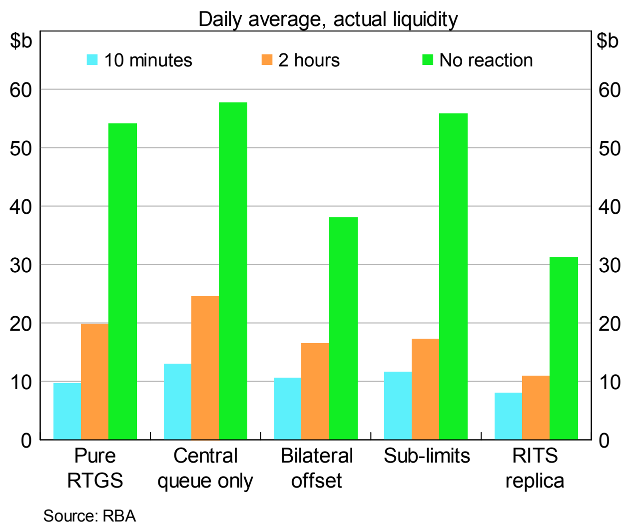
8.5. Results

8.5.1 System design and reaction times

8.5.1.1 Actual liquidity

The results confirm the mitigating effect of quicker participant reactions established in Ledrut (2007); the shorter the time taken to stop payments to the participant with the outage (ie the reaction time), the lower the average value of unsettled payments between non-stricken participants (Figure 8.3). If participants do not react, the average value of unsettled payments at the end of the day can be up to $57.6 billion, whereas the average value of unsettled payments when participants react 10 minutes after the disruption ranges from $7.9 billion to $12.9 billion, depending on the system design.

Figure 8.3 **Unsettled payments**



In line with expectations, introducing a bilateral offset algorithm reduces the systemic impact of participant operational outages. While the results suggest that introducing a central queue, in and of itself, does not mitigate the systemic effect of participant operational disruptions – in fact, the systemic impact is slightly larger – this is probably due to methodological issues.[17] Compared to the central queue-only system, average unsettled payments in the bilateral offset system decrease by between $2.3 billion and $19.7 billion, depending on the participant reaction time. Notably, a quick reaction by other participants is less important in systems with bilateral offset. Of course, the caveat to this result, which is investigated further in the following sub-section, is that participants may respond to the inclusion of a liquidity-saving algorithm by decreasing the amount of liquidity they hold, which may overstate the benefit of having such an algorithm.

As noted in the introduction, a liquidity reservation feature slows liquidity recycling, which could increase or decrease the systemic impact of a participant operational disruption. The net effect of sub-limits depends on participants' reaction time; unless participants react and lower their sub-limits, the flow of liquidity remains restricted and sub-limits do not mitigate the systemic impact of a participant operational disruption. However, because sub-limits slow the flow of liquidity into the stricken participant's account, sub-limits minimise the effect of a longer reaction time, ie for a 2 hour reaction time, the liquidity sink is $2.0 billion smaller and the value of unsettled payments is $7.4 billion less, than in the central queue-only system. The case where non-stricken participants do not drop their sub-limits is considered in Section 8.5.1.3 below.

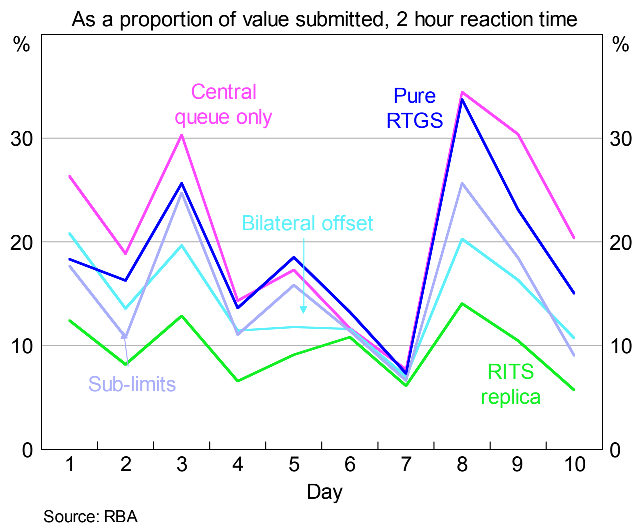
In our simulations, the systemic consequences of an operational disruption, across all the reaction times, are minimised by the combination of sub-limits and bilateral offset in the RITS replica system. Interestingly, the results suggest that – as long as participants react – combining bilateral offset and sub limits in a single system

[17] Both the method of selecting the disruption and the submission time assumptions are likely to understate the benefit of a central queue. Firstly, the method of selecting the disruption to simulate virtually eliminates the mitigating effect of queued payments; since queued payments decrease the size of the theoretical liquidity sink, the largest liquidity sink generally occurs when there are minimal queued payments from the stricken participant. Secondly, the pure RTGS system is simulated using the settlement times from the benchmark central-queue-only simulations as submission times. This is done in an attempt to capture the co-ordination of payments using internal schedulers in a pure RTGS system. However, simulating a participant disruption changes the settlement times in both systems, potentially making the systemic consequences of an operational disruption in the central-queue-only system relatively larger. Furthermore, while unstricken participants are likely to react by delaying all payments to the stricken participant, the simulator can only model the delay of unsent payment. Given the earlier submission times in the central queue systems there are likely to be more queued payments to the stricken participant, which will still be tested for settlement even after participants react.

amplifies the effect of the bilateral offset algorithm. Specifically, if participants react after 2 hours, unsettled payments in the RITS replica system are reduced by a further \$5.6 billion, on top of the \$8.1 billion reduction from introducing the bilateral offset algorithm. Similarly, for a 10 minute reaction time, the value of unsettled payments is reduced by a further \$2.6 billion on top of the \$2.3 billion reduction when bilateral offset is introduced.

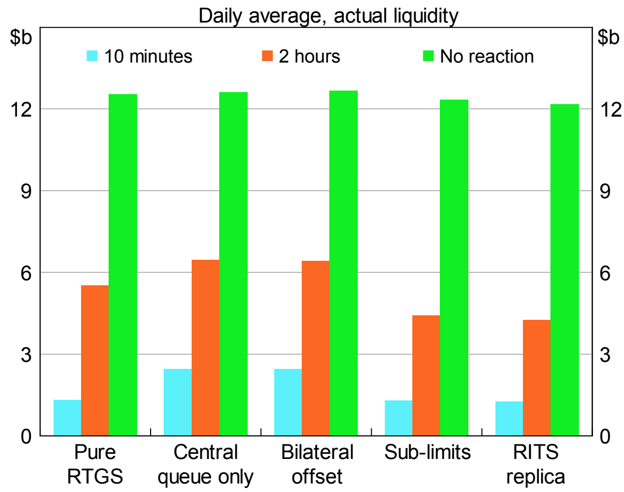
While there are sizeable inter-day variations in the value of unsettled payments, particularly in the pure RTGS system with a 2-hour reaction time where the value varies between \$7.8 billion and \$35.5 billion, this is largely related to the value submitted on a particular day. Taking the value of payments submitted (excluding payments involving the stricken participant) into account, the variability of unsettled payments in the systems with bilateral offset is significantly lower (Figure 8.4). For example, unsettled payments as a proportion of value submitted (for a 2 hour reaction time) in the RITS replica system is fairly stable at around 10 per cent.

Figure 8.4 **Unsettled payments**



As noted above, sub-limits can mitigate the systemic effects of a participant operational disruption by slowing the flow of liquidity into the liquidity sink. In systems with sub-limits, the liquidity sink is between \$1.2 billion and \$2.2 billion smaller than in the central-queue-only system, provided that participants react to the disruption (Figure 8.5). However, if participants do not stop payments to the stricken participant (only dropping sub-limits at the end of the day), sub-limits slow the development of the liquidity sink without increasing the liquidity available to settle payments between unstricken participants. As a result, when sub-limits are dropped only at the end of the day, the remaining queued payments to the stricken participant settle and the size of the liquidity sink is unchanged from systems in which there were no sub-limits. In line with expectations, bilateral offset has no major impact on the size of the liquidity sink.

Figure 8.5 **Liquidity sinks**

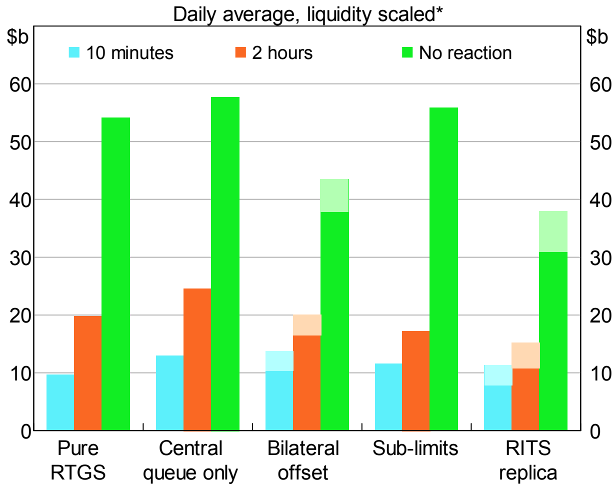


Source: RBA

8.5.1.2 Scaled liquidity

This sub-section reports the results from the simulations in which the liquidity available in the systems that include a bilateral offset algorithm was reduced to model the effect of participants responding to the inclusion of liquidity saving algorithm by reducing their holdings of liquidity. Relative to the pure RTGS scenario, liquidity is reduced by 30 per cent. In this case, unsettled payments in the bilateral offset system increase by between $3.1 billion and $5.5 billion (the lighter shaded areas in Figure 8.6). With a 10 minute reaction time, the decrease in liquidity negates the liquidity saving benefit of the bilateral offset algorithm when compared with the central-queue-only system. While the inclusion of a bilateral offset algorithm does mitigate the systemic impact of a disruption when participants react after 2 hours, the reduction in liquidity means the bilateral offset algorithm, by itself, is less effective than sub-limits.

Figure 8.6 **Unsettled payments**



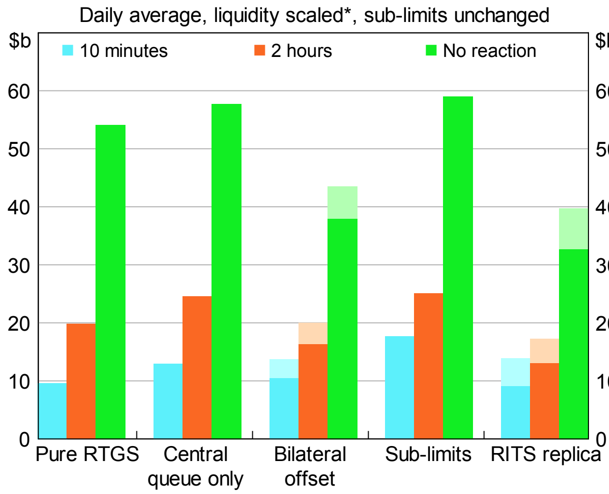
* The lighter shaded segments of the columns represent the additional unsettled payments resulting from the 30 per cent reduction in liquidity in the systems with bilateral offset.
Source: RBA

Even when participants reduce their liquidity, the RITS replica system remains the most effective system for minimising the systemic impact of a participant operational disruption. A 30 per cent reduction in liquidity causes average unsettled payments in the RITS replica system to increase by between $3.3 billion and $6.7 billion, with longer reaction times resulting in larger increases in unsettled payments.

8.5.1.3 Sub-limits maintained

In the reduced liquidity scenarios, if non-stricken participants choose to maintain their sub-limits when they react to the operational disruption, the value of unsettled payments in the sub-limit only system increases by between $3.2 billion and $7.9 billion as liquidity trapped by the sub-limits is not recycled (Figure 8.7). Similarly, as a result of maintaining sub-limits the value of unsettled payments in the RITS replica system also increases by between $1.5 and $2.8 billion. Nevertheless, unsettled payments are still lowest in the RITS replica system compared to other system designs unless participants react after 10 minutes, in which case the bilateral offset system produces a slightly better result, with \$100 million less unsettled payments compared with the RITS replica system.

Figure 8.7 **Unsettled payments**



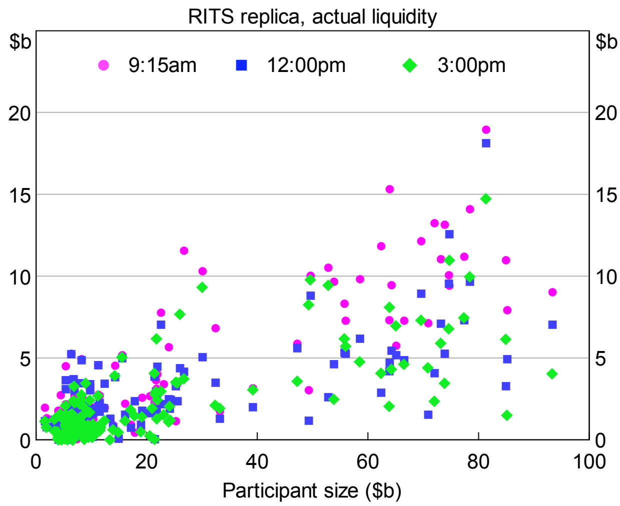
* The lighter shaded segments of the columns represent the additional unsettled payments resulting from the 30 per cent reduction in liquidity in the systems with bilateral offset.
Source: RBA

8.5.2 System design and participant size

It is generally assumed that the larger the participant (measured in terms of the value of a participant's payments and receipts) experiencing the operational disruption, the larger the systemic effects of that disruption. However, as the results show, this is a simplification, since the intraday timing of the disruption and the stricken participant's liquidity and queue management behaviour also affect the systemic impact of the disruption.

To examine this, operational disruptions at the largest 15 participants are simulated assuming a 2 hour reaction time. The results show that the impact of an operational disruption varies depending when the disruption is assumed to have occurred. Figure 8.8 shows, for the RITS replica system, the relationship between the size of the stricken participant (measured as the total value of payments submitted to the system on that day to which it was a counterparty) and the systemic impact (measured as unsettled payments) if an operational disruption occurred at 9.15am, 12.00pm or 3.00pm. In general, the value of unsettled payments is greatest when the disruption starts at the beginning of the day, with the midday disruptions generally having a slightly larger impact that the afternoon disruption. This ordering broadly holds across all system designs simulated. Since this paper models a rest-of-day disruption, this is to be expected as there are more payments yet-to-be settled earlier in the day. The relatively small difference between the midday and afternoon disruptions is most likely due to the peak in value settled in the two hours after 3.00pm (before other participants react) increasing the theoretical liquidity sink.

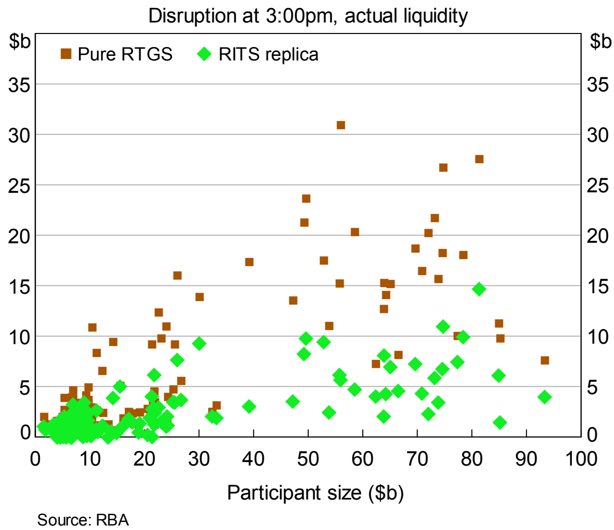
Figure 8.8 **Unsettled payments**



Source: RBA

The results also show that the inclusion of hybrid features reduces the systemic impact of an operational disruption for a participant of a given size – that is the relationship between participant size and the value of unsettled payments is flatter in hybrid systems. Figure 8.9 compares the systemic impact of operational disruptions commencing at 3.00pm in the pure RTGS and the RITS replica system across the largest 15 participants. Broadly, for a given participant size, the value of unsettled payments is lower in the RITS replica system than the pure RTGS system.

Figure 8.9 **Unsettled payments**



Some more detailed analysis of individual results underscores the importance of participants' liquidity and queue management strategies. In a number of instances where the stricken participant tends to submit payments earlier than its peers, and the operational disruption occurs later in the day, the presence of queued payments minimises the systemic impact of a disruption at that participant relative to peers that submit payments later.

8.6 Conclusion

The results of simulations conducted in this paper suggest that the systemic impact of operational disruptions of participants is generally mitigated by the inclusion of hybrid features in an RTGS system. The bilateral offset algorithm when combined with sub limits is the most effective way to mitigate the systemic consequences of an operational disruption. While the inclusion of a bilateral offset algorithm minimises the value of unsettled payments resulting from an operational disruption, the extent of this beneficial effect is reduced if participants respond by reducing their holdings of liquidity. Sub-limits can reduce the systemic impact of an operational disruption, as long as participants react to the disruption by stopping sending payments to the stricken participant and setting their sub-limits to zero to maximise the liquidity available. Unfortunately, methodological issues make it difficult to come to any firm conclusions regarding the benefits of introducing a central queue, in and of itself.

Simulated disruptions at the largest 15 participants also demonstrate that hybrid features tend to flatten the relationship between participant size and the systemic impact of a disruption at that participant.

When interpreting the results of this paper, the potential effect of endogenous behavioural responses (which is beyond the scope of this paper) need to be considered. In particular, the assumptions are likely to understate the benefits of incorporating hybrid features to the extent that then encourages earlier submission of payments. A logical extension to this work would be to incorporate expected changes in submission behaviour due to changes in system design, and the effect of these changes on the systemic impact of operational disruptions.

Appendix

Simulator Algorithms

The algorithms in the Bank of Finland simulator were modified to broadly replicate the hybrid features in RITS. In particular, we modified the bilateral offset algorithm to test all queued transactions. Our algorithm tests each queued transaction against up to a maximum of 10 offsetting transactions, starting with the first queued offsetting transaction and adding, in FIFO order, up to nine further offsetting transactions. Applying this algorithm, 27 per cent of the total value of settlements is settled via bilateral offset, compared to 25 per cent of payments in RITS which are settled via bilateral offset. This compares with around 5 to 10 per cent using the unmodified algorithm (which only tests for simultaneous settlement of transactions between the counterparties of the first queued transaction). The remaining difference between the bilateral offset algorithm constructed for the purposes of the simulation and the RITS functionality is that in the simulator payments are only tested for bilateral offset once all payments have been tested for individual settlement. In contrast, RITS tests each payment first for individual settlement then for bilateral offset (as long as the payment has been queued for at least a minute), before moving on to the next payment (Figure A8.1).

The entry, queue and bilateral offset algorithms have been modified to broadly match RITS' sub limit functionality in the simulator. Based on a payment's status these algorithms adjust the amount of liquidity available to settle that payment based on the sub-limit data, which was entered using the bilateral limits input table. Given the inability to allow for changes to payment status in the simulations, as well as a lack of data on precisely what time these changes occurred, the rules of thumb used to determine a payment's status and the submission time are as shown in Table A8.1. These are based on when the status was most likely to have changed.[18]

[18] Payment status is discussed in greater detail in Section 8.3.

Figure A8.1 **RITS settlement tests**

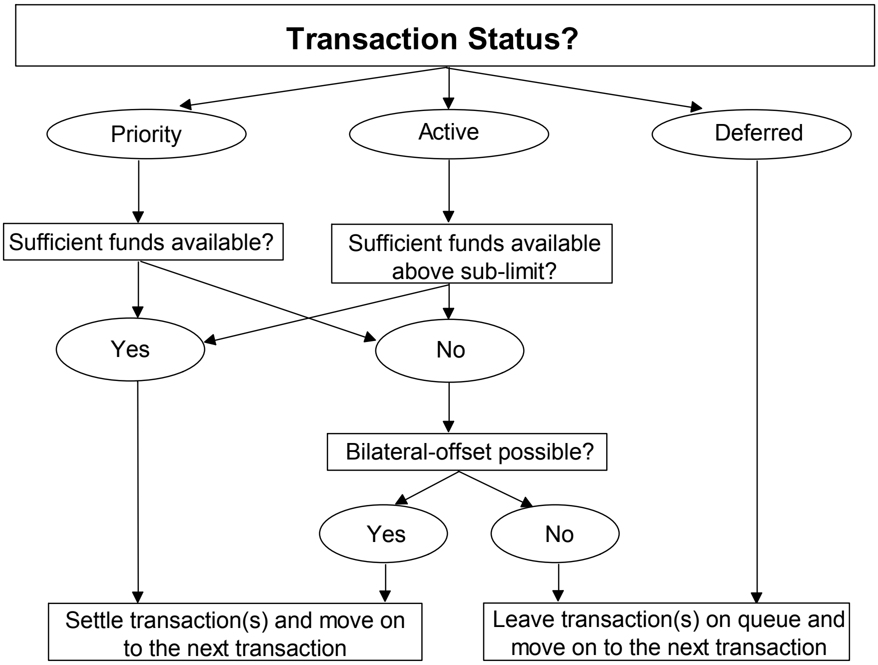


Table A8.1 **Payment status and submission times**

Status when submitted to RITS	Status when settled in RITS	Status when submitted to the simulator	Time when submitted to the simulator
Deferred	Active Priority	Active Priority	Settlement time in RITS Settlement time in RITS
Active	Active Priority	Active Priority	Submission time to RITS Submission time to RITS
Priority	Active Priority	Priority Priority	Submission time to RITS Submission time to RITS

Note: In the pure RTGS system with unlimited liquidity, all payments are submitted to the simulator at the time they were settled in RITS and payment status is irrelevant.

References

Andersen, K S – Madsen, I (2009) **A quantitative assessment of international best practice for business continuity arrangements in payment systems.** In Leinonen, H (ed), Simulation Analyses and Stress Testing of Payments Networks, Bank of Finland, 17–58.

Bank for International Settlements (2005) **New developments in large-value payment systems.** Committee on Payment and Settlement Systems Publications No. 67.

Bech, M L – Preisg, C – Soramäki, K (2008) **Global Trends in Large-Value Payments.** Federal Reserve Bank of New York Economic Policy Review, 14(2), 59–81.

Bedford, P – Millard, S – Yang, J (2005) **Analysing the impact of operational incidents in large-value payment systems: a simulation approach.** In Leinonen, H (ed), Liquidity, Risks and Speed in Payment and Settlement Systems – A Simulation Approach, Bank of Finland, 249–276.

Ercevik, K – Jackson, J (2009) **Simulating the impact of a hybrid design on the efficiency of large-value payment systems.** In Leinonen, H (ed), Simulation Analyses and Stress Testing of Payments Networks, Bank of Finland, 188–224.

Gallagher, P – Gauntlett, J – Sunner, D (2010) **Real-time Gross Settlement in Australia.** RBA Bulletin, September 2010, 61–70.

Galbiati, M – Soramäki, K (2010) **Liquidity-saving Mechanisms and Bank Behaviour.** Bank of England Working Paper No 400.

Glaser, M – Haene, P (2009) **Liquidity Effects of a Participant-level Operational Disruption in SIC.** In Leinonen, H (ed), Simulation Analyses and Stress Testing of Payments Networks, Bank of Finland, 59–82.

Johnson, K – McAndrews, J J – Soramäki, K (2005) **Economizing on Liquidity with Deferred Settlement Mechanisms.** In Leinonen, H (ed), Liquidity, Risks and Speed in Payment and Settlement Systems – A Simulation Approach, Bank of Finland, 179–216.

Ledrut, E (2007) **How Can Banks Control Their Exposure to a Failing Participant?** In Leinonen, H (ed), Simulation Studies of Liquidity Needs, Risks and Efficiencies in Payment Networks, Bank of Finland, 227–252.

Lubloy, A – Tanai, E (2009) **Operational disruption and the Hungarian real time gross settlement system (VIBER).** In Leinonen, H (ed), Simulation Analyses and Stress Testing of Payments Networks, Bank of Finland, 17–58.

Manning, M – Nier, E – Schanzk, J (2009) **The Economics of Large-value Payments and Settlement: Theory and Policy Issues for Central Banks.** Oxford University Press.

Martin, A – McAndrews, J J (2008) **An economic analysis of liquidity-saving mechanisms.** Federal Reserve Bank of New York Economic Policy Review, 14(2), 25–39.

Mazars, E – Woelfel, G (2005) **Analysis, by simulation, of the impact of a technical default of a payment system participant.** In Leinonen, H (ed), Liquidity, Risks and Speed in Payment and Settlement Systems – A Simulation Approach, Bank of Finland, 297–320.

Merrouche, O – Schanz, J (2009) **Banks' intraday liquidity management during operational disruptions: theory and evidence from the UK payment system.** Bank of England Working Paper No 370.

Schmitz, W S – Puhr, C (2007) **Risk concentration, network structure and contagion in the Austrian Real Time Interbank Settlement System.** In Leinonen, H (ed), Simulation Studies of Liquidity Needs, Risks and Efficiencies in Payment Networks, Bank of Finland, 183–226.

Chapter 9

Systemically important participants in the ReGIS payment system

*Horatiu Lovin**

* The author is grateful to Bank of Finland for providing technical support and expertise in order to implement in Romania the BoF-PSS2 simulator, an instrument developed to better understand risks faced by payment and settlement systems.

9 Systemically important participants in the ReGIS[1] payment system

Abstract

The aim of this study is to identify those credit institutions that may jeopardize the stability of the payment system and to assess the overall impact on the payment system of a systemically important participant trigging a severe disruption. The study should also be of benefit for oversight activity, eg in the event that systemically important participants are asked to improve their intraday liquidity management or even to build up capital buffers.

9.1 Introduction

According to the European Central Bank, an institution is systemically important if it meets the following conditions: (i) size, absolute or relative; (ii) interconnectedness, linkages with the other participants; (iii) substitutability, the extent to which other participants can provide the same services in the event of failure. The first two criteria are initially considered in assessing participants' systemic importance, while substitutability is tested using a stress scenario to validate preliminary results. Whereas the ongoing global financial turmoil has revealed the great importance of highly interconnected institutions in assessing systemic and contagion risk, our results point to a mild impacts on payment system for highly interconnected but smaller participants. However, our focus here is on participants that fulfilled at least one of the first two criteria, as systemic risk involves a great deal of uncertainty and all individual criteria must be carefully tested.

The main tool used in this study is the BoF-PSS2 Simulator, developed by Bank of Finland. The software replicates payment system functioning and enables sophisticated research on payment systems.

---

[1] Romanian RTGS large value payment system.

9.2 Data

The data available for the study cover the first five months of 2010 (between January and May). We use participants' daily balances and transactions carried out in the ReGIS payment system, as well as money market transactions between participants. There are 41 participant credit institutions in the payment system during the selected time period.

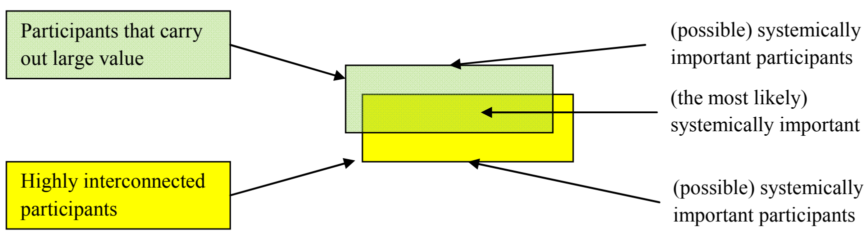
9.3 Systemically important participants

9.3.1 Size criterion

A systemically important participant within ReGIS payment system carries out large value transactions and is highly interconnected with the other participants. The participants that fulfil both criteria are the most likely ones to be considered systemically important. However, due to the complex nature of systemic risk, the participants that fulfil only one criterion are also considered for systemic-importance testing.

The definition used here for systemically important participants is that of the European Central Bank, as stated in the Financial Stability Review of June 2010. When a liquidity shortage impairs a participant's ability to settle payments, the entire system may be at risk. Therefore, the stress test scenario will involve a direct impact on selected participants (based on the previous two criteria), followed by a contagion effect across the entire ReGIS payment system (Box 9.1).

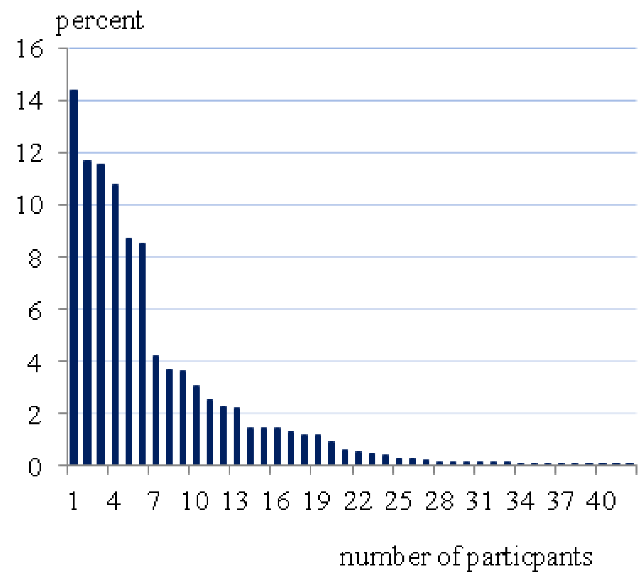
Box 9.1 **Systemically important participants**



The substitutability criteria is assessed by running stress test scenario for large and/or interconnected participants

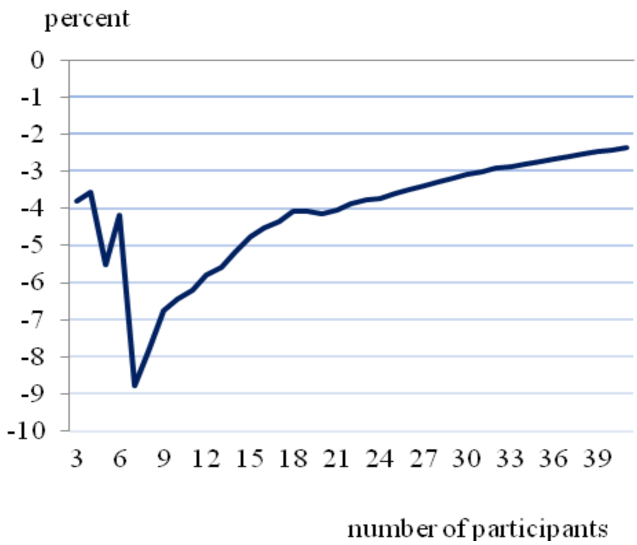
Starting with the first criteria (size), we sorted the participants based on total submitted and received payment values (Chart 9.1). Moreover, a threshold was chosen in order to split the largest participants from the rest of the sample. We built the cumulative average size of the sorted participants and selected the threshold corresponding to the participant with the strongest marginal impact (Chart 9.2). This methodology provides a simple cut-off point within the participant-sample, sorted by relative size. The point selection represents the trade off between obtaining a reasonable number of large participants and looking for an abrupt change within the sorted sample.

Chart 9.1 **Market share for each participant in ReGIS payment system in terms of settled payments value**



Source: NBR (National Bank of Romania)

Chart 9.2 **The marginal impact on cumulative average of participants' market share**



Source: NBR

The marginal impact on the cumulative average of participants' market share was calculated as follows: (i) first, the cumulative average of participants' market share was computed for the participants in descending order (arithmetic average for the first two participants, then for the first three participants etc, ending with the arithmetic average for all 41 participants; (ii) second, the marginal impact on the cumulative average of participants' market share was calculated as the impact of each new participant was inserted into the average sample, according to the formula bellow

$$\text{impact of participant i} = \frac{\text{the average of the first i participants} - \text{the average of the first i} - 1 \text{ participants}}{\text{the average of the first i} - 1 \text{ participants}}$$

(iii) the strongest marginal impact was selected as the threshold for splitting the sample into large and small participants.

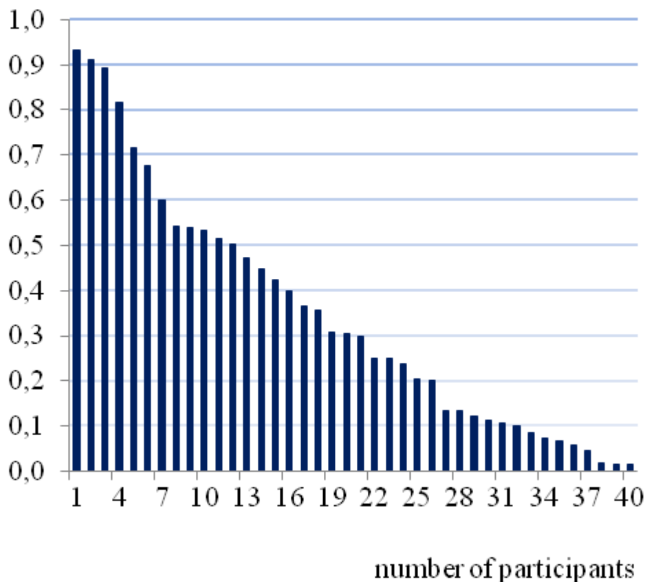
Based on Chart 9.2, participant number 7 has the strongest impact on the cumulative average of participants' market share; hence the first 6 participants are designated as large (see Chart 9.1).

9.3.2 Connectivity criterion

Regarding the second criteria (interconnectedness), we apply the network analysis and build a connectivity index that weights equally the total volume of interbank deposits and the average number of daily connections (number of banks interacted with). Based on the connectivity index for each participant, we applied the same method as for the size criterion, sorting the participants and then setting the threshold. We thus obtained a sample containing the most interconnected participants.

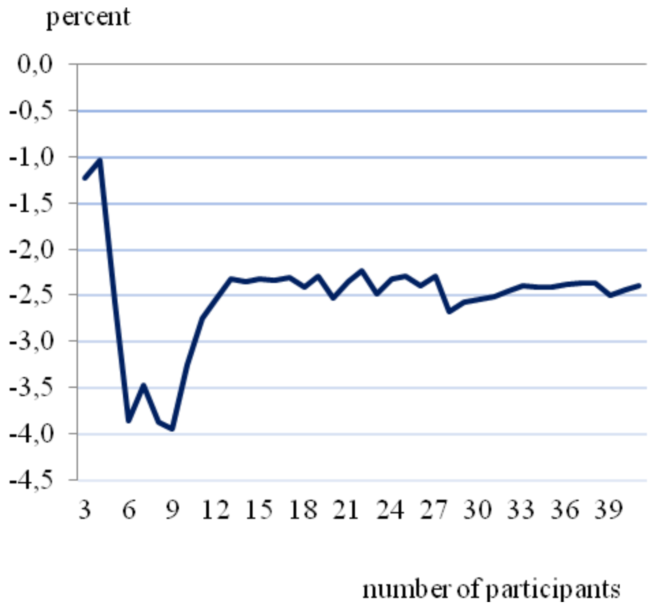
Because the connectivity index is computed as the average of two variables (total value of interbank deposits and daily average number of connections), the threshold for splitting the sample into highly versus less interconnected participants is not as obvious as for the size criterion. The descending slope is smoother (Chart 9.3) and the marginal impact on cumulative average of participants' connectivity index is lower (Chart 9.4).

Chart 9.3 **Connectivity index for each participant in ReGIS payment system**



Source: NBR

Chart 9.4 **The marginal impact on cumulative average of participants' connectivity index**



Source: NBR

Based on Chart 9.4, participant number 8 has the strongest impact on the cumulative average of the participants' connectivity index; therefore, the first 7 participant are designated as highly interconnected (see Chart 9.1).

There are 6 participants that passed through the size filter (the largest 6) and 7 participants that met the interconnectedness criterion (the 7 most interconnected). However, these two samples are not very homogeneous because the filters were applied so as to remove participants with little impact on the payment system. Taking into account the heterogeneity of the two samples, we ranked the selected participants (6 large and 7 highly interconnected) and obtained 6 large participants (1 large, 3 medium, 2 low) and 7 highly interconnected participants (3 high, 2 medium, 2 low). These two samples include 9 participants in all, 4 of which participants are both large and highly interconnected.

9.3.3 Participants' behaviour

Credit institutions place payment orders during the trading session according to customer orders and their own needs and consistent with the available funds. Because there is uncertainty about the times and amounts of liquidity coming via other participants' payments, all participants have an incentive to delay their payment submissions in order to minimize their liquidity needs. Credit institutions must borrow money by the end of trading session if they lack sufficient liquidity to settle their submitted payment orders and borrowing money is costly. Therefore, participants prefer to wait for incoming payments before submitting their own payment orders if the level of available funds is low. Nevertheless, participants face reputational risk if they excessively delay payments; hence they are looking for an optimal solution that balances costs and reputation.

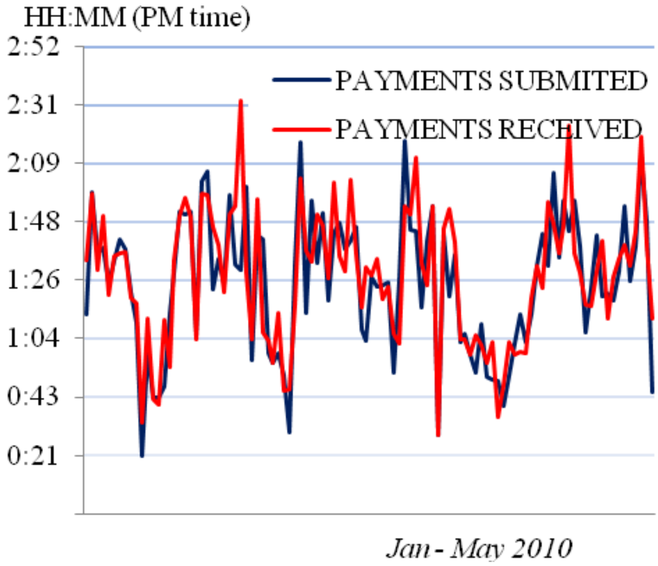
Participants' behaviour impacts liquidity flows within the payment system. Payment system stability is enhanced by participants that inject liquidity into the system, which help to ensure smooth system functioning and prevent excessive build-up of liquidity risk inside the system.

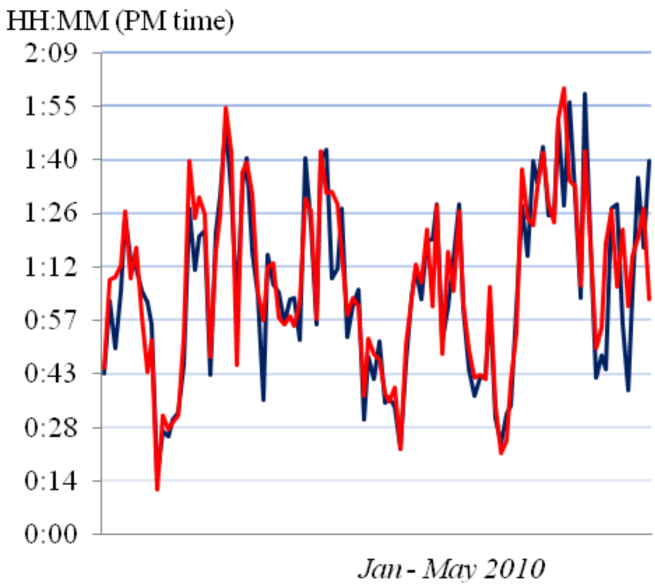
Charts 9.5 and 9.6 plot the weighted average time of payments submitted and received for the two largest and two most interconnected participants. The large value transactions carried out by the largest participants tend to occur in the second half of the trading session, so that the weighted average time for the largest participants is higher than for the highest interconnected ones which submit and receive payments earlier in the day.

The spread between the weighted average time for submitted and received payments is very narrow for all four participants, emphasizing the similar behaviour across participants (Charts 9.5 and 9.6). The fact that large value transactions are carried out in the second half of trading session puts some pressure on the available liquidity. However, participants can settle large value payments within a short period of time due to the large amounts of usable resources and timing benefits.

It is interesting that the largest participants are not the same as the most interconnected ones, which is further evidence of the complex nature of systemic importance. Both criteria are equally important in assessing participants' systemic importance.

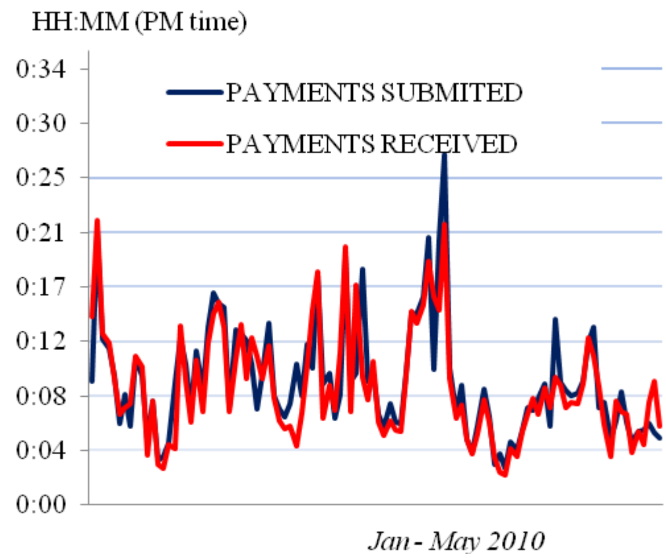
Chart 9.5 **Daily average transaction time for the largest two participants in ReGIS payment system**

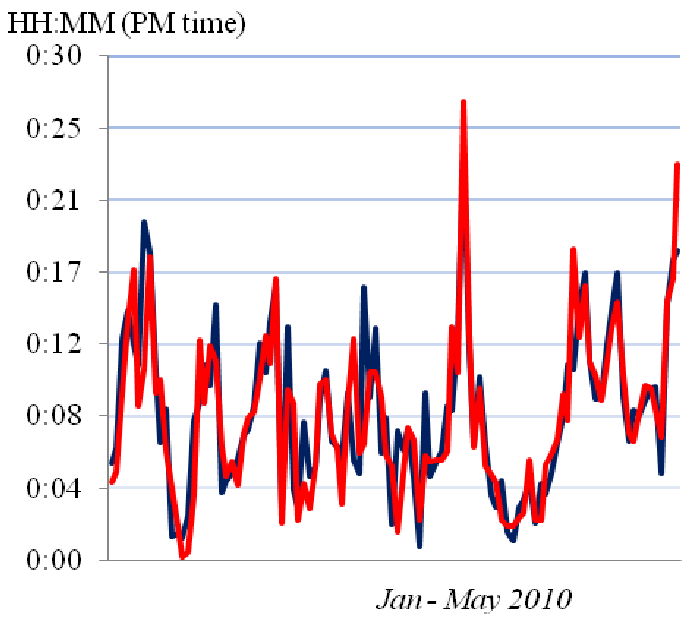




Source: NBR

Chart 9.6 **Daily average transaction time for the highest two interconnected participants in ReGIS payment system**





Source: NBR

9.3.4 Substitutability criterion (stress test)

The substitutability criteria had been tested by running a stress scenario, based on following assumptions: (i) a participant that fulfils at least one of the first two criteria (size or connectivity) experiences a severe liquidity shortage and his balance account drops to zero at the beginning of the day (session); (ii) the only liquidity available for the participant to settle the submitted payment orders is the incoming payments from other participants; if there is not enough liquidity in his account, the payment order is queued; (iii) none of the participants (including the one with the severe liquidity shortage) changes the times of payment submission. We use the BoF-PSS2 Simulator to run the scenario for each possible systemically important participant.

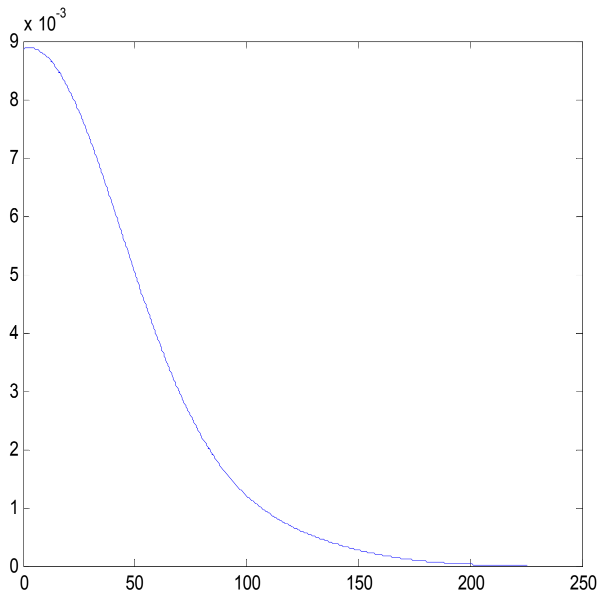
The losses triggered by a major liquidity disruption within the payment system can be quantified in different ways. We mention only a few of them: (i) maximum daily queues value; (ii) the value of unsettled payment orders at the end of the day; (iii) total costs for liquidity deficit participants to raise funds from the money market; (iv) the overall impact on the real economy. The maximum daily queues value is a measure of the intraday liquidity deficit among payment system participants at the aggregate level. Still, the overall impact on the real economy requires a distinct analysis. The total costs for participants with liquidity deficits that need funding from the money market may not be high if they own enough high quality assets that qualify for transactions with the central bank.

We assume that a systemically important participant cannot be substituted, if the other participants do not have enough excess liquidity for settling payments and offsetting the liquidity shortfall of the systemically important participant. The results reveal a stronger impact (a higher value for maximum queued payments) for the participants that meet both criteria (large and highly interconnected), compared to those participants that are either large or highly interconnected (see Appendix). However, none of the participants that fulfil at least one criterion (size or connectivity) should be removed from the top monitored list. Even a small shock (in terms of behavioural incentives) may cause severe damage, and pure contagion can play a major role in spreading risks across the system.

9.3.5 Contagion

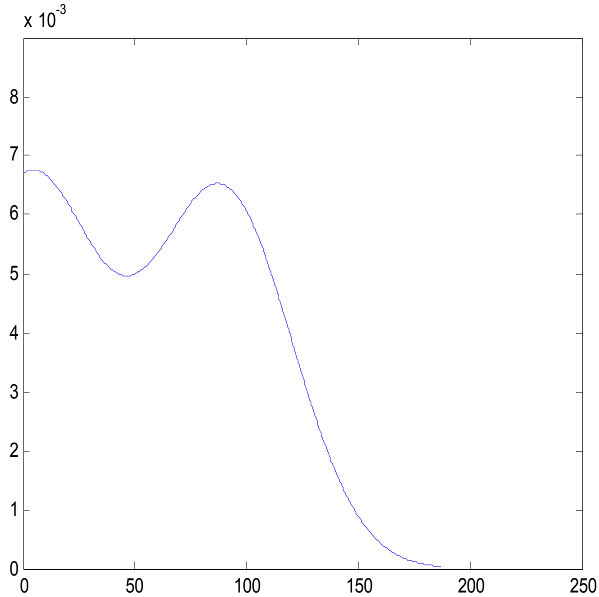
The overall queued payments revealed by stress test results are submitted ether by the participant experiencing a sudden liquidity shock or by other participants that suffer fund deficits due to a liquidity flow disruption triggered by the aforementioned participant. The queued payments submitted by the other participants reflect contagion risk materialisation. In light of the stress test results, it seems that the highly interconnected participants trigger stronger contagion compare to the large participants. This is the case because some of the highly interconnected participants provide liquidity to the banking sector (their impact is specific rather than overall), while the large participants provide liquidity to the entire payment system, thus smoothing the flow of funds inside the system. The results for both the largest and most interconnected participants can be observed from charts 9.7 and 9.8.The kernel distribution for the highest interconnected participant has a fatter tail compared to the largest participant, suggesting larger contagion losses triggered by the former participant.

Chart 9.7 **Contagion triggered by the largest participant (kernel distribution for the queue values)**



Source: NBR

Chart 9.8 **Contagion triggered by the highest interconnected participant (kernel distribution for the queue values)**



Source: NBR

9.4 Conclusions

Systemically important participants in the ReGIS payment system can jeopardize its stability and pose a threat to the real economy via liquidity flow disruptions. We observe the ECB criteria for assessing participants' systemic importance: size, connectivity and substitutability.

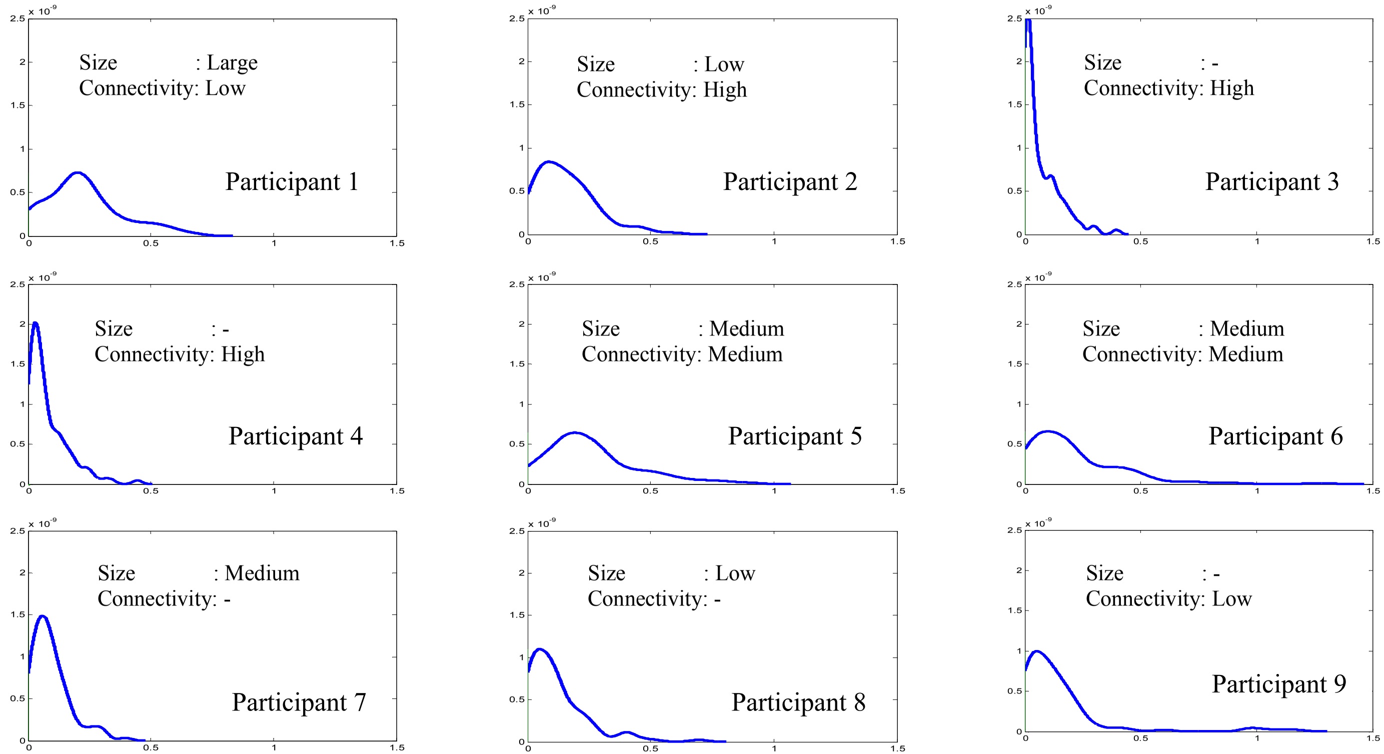
Our findings emphasize that participants both large and highly interconnected cannot be substituted by the other participants and thus, as they meet all three criteria, they are considered systemically important participants. Participants that are either large or highly interconnected can be substituted for to a large extent by other participants, so that they are less likely to be considered systemically important.

The large participants pose less contagion risk for payment system, in the event of a liquidity shortfall, than do the highly interconnected ones; but the overall impact on the system is stronger for the large participants. The highly interconnected participants provide liquidity to the banking sector (their impact is specific rather than overall), while the large participants provide liquidity to the entire payment system, smoothing the flow of funds inside the system.

Finally, we underline the fact that, due to the high degree of dependence of our results on the input data, we do not exclude the possibility that by extending the period analysed and/or observing the payment system functioning in a different economic and financial environment, the results might change.

Appendix

Chart 9.9. **Stress test results: the kernel distributions for daily maximum queue values when a participant fails**



Source: NBR

Chapter 10

Network dynamics of TOP payments*

*Marc Pröpper** – Iman van Lelyveld** − Ronald Heijmans****

* This paper is a revised version of `Towards a network description of interbank payment flows'. It reflects the opinions of the individual authors and does not necessarily reflect the views of De Nederlandsche Bank. The authors would like to thank Hans Brits, Maarten Gelderman, Walter Stortelder and Kimmo Soramäki for their useful comments, and Gwan Tjia, Richard Heuver and Odette Verlee for their statistical support.

** De Nederlandsche Bank, Supervision Policy Division, P.O. Box 98, 1000 AB Amsterdam, The Netherlands. Email: marc.propper@dnb.nl and iman.van.lelyveld@dnb.nl.

*** De Nederlandsche Bank, Payments Policy Division. Email: ronald.heijmans@dnb.nl.

10 Network dynamics of TOP payments

Abstract

We present an application of network theory to the Dutch payment system with specific attention to systemic stability. The network nodes comprise of banks active in the Netherlands where links between the nodes are established by payments. Traditional measures such as transactions and values first show payments are relatively well behaved through time. Analysis of the properties of prominent network measures over time shows that network characteristics become clear in the early phase of network formation of about one hour and slower development afterwards. The payment network is small in terms of actual nodes and links, compact in terms of path length and eccentricity and sparse in terms of connectivity for all time periods. In the long run a mere 12% of the possible number of interbank connections is ever used and banks are on average only two steps apart. Relations in the network tend to be reciprocal. Our results also indicate that the network is susceptible to directed attacks. In a final section we show the effect of the start of the financial crisis on the network structure including the effects of the migration to TARGET2.

10.1 Introduction

From the late 1990s the study of the topological structure of random networks has gained momentum. Empirical observations from large and rapidly evolving networks like the World Wide Web (Albert, 1999), the Internet (Faloutsos et al, 1999), and journal publishing economists (Goyal et al, 2006) brought to light a surprising compactness ('small world phenomenon') and relatively many highly connected networks elements (Dorogovtsev and Mendes, 2003). These findings have shifted attention away from classical, static networks (Erdõs and Rényi, 1959) towards growing networks.[1] An important property of the latter is their robustness against random failures. At the same time, however, they are vulnerable against directed attacks (Albert et al, 1999). Terrorists, interested in creating the largest impact, would therefore rather attack a more central node in a scale free network than in a random network: the impact would be much larger.

The ideas of network theory can be applied to the field of economics, for example to study the risk of widespread propagation of financial distress (*systemic risk*). There is a vast literature analysing the interactions of various financial markets such as equity and bond markets.[2] A small but growing literature examines the riskiness of interbank markets, where banks exchange relatively short term and largely unsecured funds.[3] These papers, however, do not focus on the network topology of the markets. Inaoka et al (2004) and Soramäki, et al (2007) have started to describe large value financial payment systems (in Japan and the US, respectively) from a network perspective. Another example is Bech and Garratt (2006) which analyses the effects of a wide-scale disruption on the functioning of the interbank payment system. Our paper builds on this literature, and adds to it by illustrating (1) the influence of the chosen time frame on the properties of the payment network, (2) the central role of highly connected banks in the functioning of the payment network. We also show that the beginning of the crisis, which started in the summer of 2007, seems to have led to change in network structure although this analysis is hindered by the migration from TARGET1 to TARGET2.

Importantly, in contrast to for instance analyses of interbank exposures (eg Boss et al, 2004, or van Lelyveld and Liedorp, 2006), payments networks are by definition short lived: as soon as the payment is settled, the visible, recorded connection between banks disappears. This affects our understanding of what constitutes a network. Here we show that the *time frame* used to compute the

[1] The former, equilibrium random networks, have Poisson degree distributions (the degree of a node is the number of its links). The latter, non-equilibrium random networks, may under the right conditions result in fat-tailed, scale-free degree distributions close to a power law. This is the case when they are governed by (a linear kind of) preferential attachment which means that new network elements are more likely to attach themselves to elements that are already highly connected (Barabási and Albert, 1999).

[2] See Pericoli and Sbracia (2003) for an overview.

[3] See Allen and Gale (2000) for a theoretical characterisation of these markets and van Lelyveld and Liedorp (2006) and the references therein for an empirical analysis. The general finding is that interbank markets are from a systemic stability point of view relatively safe.

network measures materially affects the outcomes. A proper understanding of the evolution of the network is important from a risk management perspective. The specific vulnerabilities – and thus the optimal response – change over time. Ultimately, the purpose is to get an understanding of the level of stability or, alternatively, vulnerability of the system to random or directed failures and to systemic risk.

Network theory equips us with promising tools. The questions we will answer are tackled by first studying the (time development of the) network structure of the payment system in terms of commonly used network properties like the size of the payment network, the connectivity between banks, distances in the network, the distributions of connections between banks and network correlations. This for example allows us to take a first, indirect peek at the risks that the system faces, by assessing the importance to the network of the most highly connected banks. In addition, we present the effect of the start of the financial crisis on the payment network.

The set-up of our paper is straightforward. We start with a short description of the institutional detail of the Dutch large value payment system (TOP), the technical details of the data set and an international comparison of aggregate key figures. Next, we discuss the intraday behaviour of the system. Then we investigate whether there are structural imbalances (between individual – or groups of – participants) in the system. Such imbalances are revealed by persistent payment flows. After this examination of the basic properties of our data we examine the build up of the network over time. First we analyse the development over time of commonly used network measures in the literature. Second, we analyse the vulnerability, or, alternatively formulated, systemic stability, of the system. We report the impact on the network structure and on the key system figures of removing the ten most highly connected participants in the data set. In addition, we analyse whether the recent 'sub prime' crisis in credit markets has affected the network properties of the payment system. We end with the conclusions.

10.2 The dutch payment system

Previously the Dutch large value payment system (TOP) formed part of the European system for euro-denominated payments, TARGET.[4] TOP was restricted to a limited set of participants, mainly banks. Connections to participants in other TARGET countries takes place through TARGET. As of 18 February 2008, settlement has migrated to TARGET2. Technically, this is a centralised system, but legally it is a decentralised system in which each country designates its own component system. We will return to the operation of TARGET in our section on the effect of the crisis.

For the system to function properly it is essential that participants have sufficient funds so that payments can be made without delay. Intraday credit provided by DNB (secured by collateral) facilitates a smooth functioning of the payment system and prevents gridlock.[5] In the Netherlands, commercial banks permanently hold (pledged) collateral at the central bank, generally at a relatively stable level during the year.[6]

Regular opening hours for TOP were from 7h until 18h and during these hours the payment system processes all transaction types. In addition, there was an evening settlement period from 19:30h to 22h used for settling ancillary system batches and not for standard domestic transactions and cross border (TARGET) payments. Incidentally, the latter two types of transactions made up for more than 80% of the value transferred (Oord and Lin, 2005).

For our analysis of the 'normal' period, we analyse a data set consisting of one year of transaction data from the Dutch large value payment system, running from June 2005 to May 2006 (257 business days).[7] Transactions carried out during evening settlement are generally excluded, except for calculations of net value transfer. No standard domestic, domestic correspondent bank and cross border transactions through TARGET are carried out during evening settlement. We use the settlement time rather than the moment a transaction is entered into the system in our analysis. Participants with

---

[4] Trans-European Automated Real-time Gross settlement Express Transfer system.

[5] See Ledrut (2006) for a discussion of the optimal provision of intraday liquidity.

[6] In addition to maintaining a collateral pool, it is possible to place collateral through repo transactions. When the credit balance becomes insufficient, collateral is brought in and the balance is raised, usually in the morning. At the end of the day, the transaction is reversed.

[7] Processing of data has been done in Java by extending graph data structures from Goodrich and Tamassia (2006).

more than one account are consolidated. Payments between two accounts of a single participant are therefore not included in the analysis. Also, due to the limitations of the dataset, cross border transactions are not analysed on a participant, but on a country level. In short, we analyse a network of participants, not of accounts, and some participants are countries rather than banks.

Table 10.1 shows daily averages on numbers of participants, transaction volumes, values transferred and (average) transaction values for the Top (NL), TARGET (EU), CHAPS (UK) and Fedwire (US) payment systems. The TOP figures are presented with and without evening settlement.[8] They include incoming and outgoing cross border payments through TARGET. The numbers show that TARGET and Fedwire are both large payment systems of the same order of magnitude. The Dutch domestic system is clearly smaller; only the average transaction value is relatively high.

Table 10.1 **Key figures on daily payment characteristics Top (NL), TARGET (EU), CHAPS (UK) and Fedwire (US)**

	Top (without / with evening settlement	**TARGET**	**CHAPS**	**Fedwire**
Measurement period	6/2005–5/2006	2005	2005	2005
Participants	155[9]	10,197	NA	6,819
of which direct participants	100	1,126	15	NA
Transactions (x 1000)	15.1 / 18.1	312	116	519
Value (in billion EUR)	151 / 173	1,987	297	1,634
Transaction value (in million EUR)	9.9 / 9.5[10]	6.4	2.6	3.1

Source: Top (DNB), TARGET (ECB bluebook), CHAPS and Fedwire BIS (2007).

[8] Numbers including evening settlement are relevant for section 'Net value transferred'; numbers without evening settlement are relevant for intraday payment behaviour discussed in section 'Intraday dynamics'.

[9] The number of active participants in the measurement period amounts to 129 (or 131 with evening settlement).

[10] All payments within a second from and to the same participant are aggregated. When every payment is treated separately, the average value decreases to approximately EUR 7.5 million. In case the incoming cross border payments are excluded the average payment value is EUR 6.5 million.

10.3 Traditional characteristics

Now we first turn to an examination of 'traditional' characteristics of the payment system. The network measures discussed later will not render these traditional measures obsolete: they are complementary measures. We will first look into the intraday dynamics. Relatively stable dynamics are important when considering different time frames for network measures later in this study. Secondly we will analyse the net payment position of participants over different horizons. This should tell us whether there are persistent net payers or receivers and thereby indicate if directions in payment links between participants matter when studying network properties. We will also examine the role of the three biggest banks in the Netherlands. Market concentration is high and therefore the behaviour of the large banks is an important determinant of the overall market structure.

10.3.1 Intraday dynamics

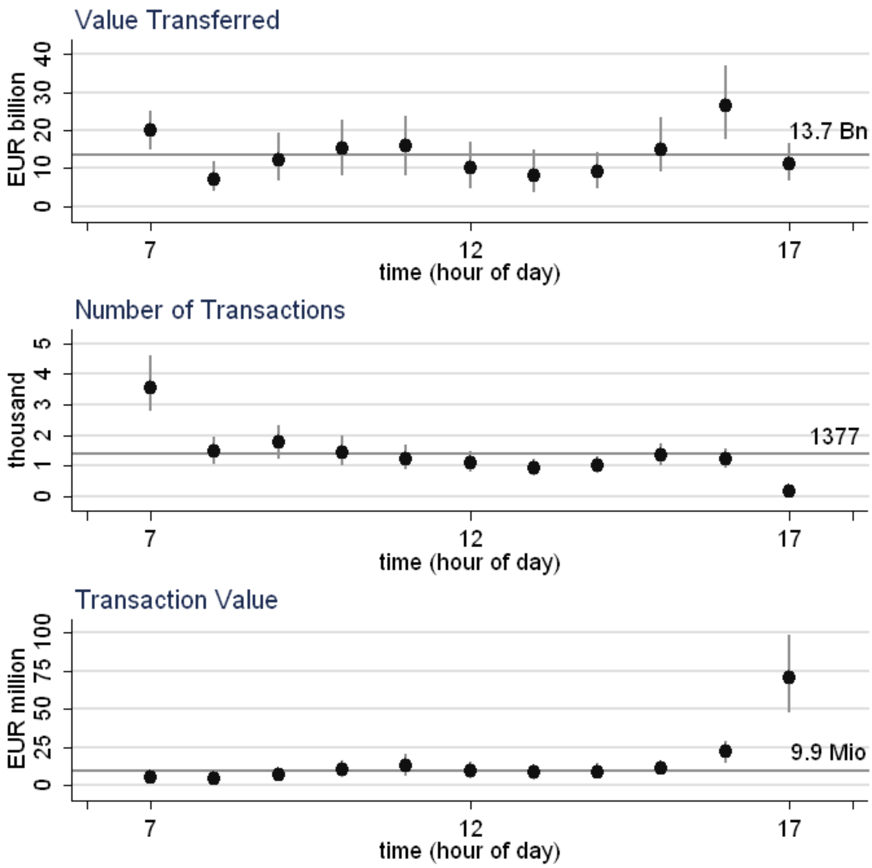
Figure 10.1 displays for each business hour the average value transferred, the number of transactions processed and the transaction value. The first pane shows Dutch banks are willing to pay early in the day: *value transferred* peaks (EUR 20.0 billion) during the first business hour.[11] This is due to payments entered the day before. Numbers strongly increase to an all day high (EUR 26.6 billion) between 16h and 17h. Some of this activity is the result of banks that need to level their balances as a result of the intraday credit used or to fulfil their cash reserve requirements. These payments are usually few in number but relatively large in value. In the last business hour, from 17h to 18h, only transactions between banks are processed (no retail orders), but most banks usually do not wait until the last hour to close their balance of the day, and finish before 17h. Therefore, value transferred slumps in the last hour. Closer inspection of the distributions around the means revealed they are fairly symmetrical. This is also the case for the distributions of the number of transactions and the transaction values.

The second pane, showing the *number of transactions*, illustrates that on average about 3,500 transactions are processed in the first hour, almost one every second. The 5% and 95% percentile values

[11] Against an average value transferred per hour (day) of EUR 13.7 (151) billion.

range from 2,800 to 4,600. The rest of the day transactions take place in smaller numbers (between 900 and 1,750 transactions) along a somewhat downward sloping trend against time that abruptly drops to very low numbers in the last hour (5% percentile ≈ 200). The average number of transactions per hour and per day respectively amount to 1,377 and 15,148. The distribution range is significant and comparable through time (between 45% to 65% of the average value for the specific hour).

Figure 10.1 **Average value transferred, number of transactions processed and transaction value during regular opening hours**



Note: the averages for a particular hour (over all of the 257 business days) are denoted by a dot; the bars run from the 5% low to the 95% high percentiles of the observations. The horizontal lines depict daily averages over regular opening hours.

Finally, the third pane shows the *transaction value*. The average transaction value during the day amounts to EUR 9.9 million but in the last two hours of business it increases strongly to respectively EUR 22 million and EUR 71 million. The last hour, however, hardly contributes to the overall average due to the small number of transactions.

Overall we observed payment characteristics do change during the day. Payment behaviour in the beginning and the end of the day differ most from the rest of the day. In the morning many payments seem to be driven by 'good customer' behaviour to pay early while payments during the last hour very likely reflect liquidity decisions mainly. We will not further distinguish between hours of the day here.

Further on, when determining the time development of network properties, the observed variability of network properties can partly be explained by the payment patterns during the day shown here.

10.4 Network measures

10.4.1 Introduction

The previous sections have described the payment system from a traditional perspective in terms of transactions processed and values transferred. This has given insight in the behaviour of individual participants and of the system as a whole. Now the perspective will shift towards describing the payment system in terms of its *network* properties.

A network (or graph in mathematics) is a set of connections (links) between pairs of objects (nodes). Examples of real-life networks are numerous and include social networks, communications networks, transportation networks, biological networks, the World Wide Web, the Internet and financial networks. In a payments network, like the one studied here, the participants form the nodes and transactions establish links between the nodes. Within each time period considered, a link between two nodes is created by the first transaction between them. Subsequent transactions add weight to the link in terms of the number of transactions processed and the additional value transferred over this link.[12] Every pair of nodes can be connected by two opposing links since the individual transactions contain a clear

[12] In this paper link weights are not taken into account in determining network properties (to prevent subjectivity) and all links are thus considered equivalent.

direction from payer to receiver, ie the network is a directed network.[13] Finally, a path is an alternating sequence of nodes and links such that each link is incident to its predecessor and successor nodes. A path can be directed (along directed links) or undirected (along undirected links).

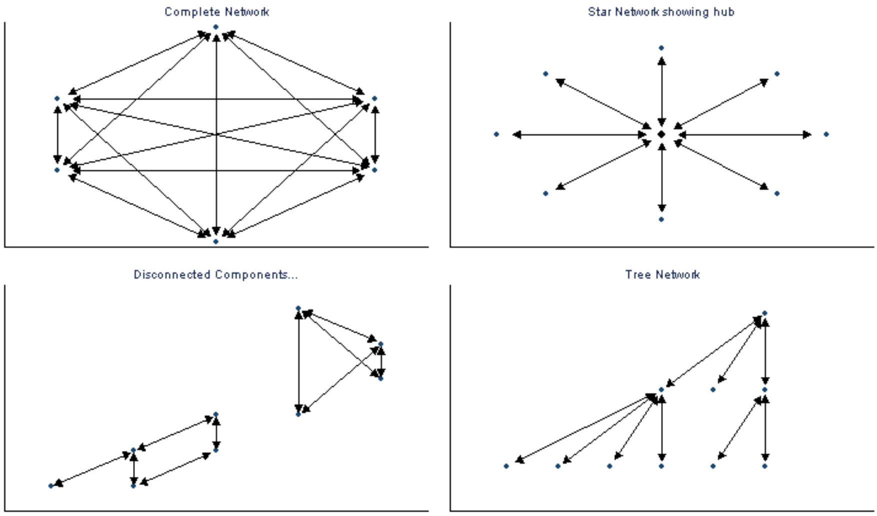
For illustrative purposes graphical representations of some basic (undirected) network types are shown in Figure 10.2. These include a complete network, a star network, a tree network and a network with two disconnected components. In a complete network all nodes are connected to all other nodes by a link. A star network is a network in which the nodes connect to a central node called the hub. In a tree network all nodes are connected by exactly one path (no loops or cycles). In a network component all nodes are connected by at least one path. A network is connected if it consists of a single component; if a network is not fully connected it consists of two or more components.

For an introduction into the theory of random networks and the treatment of real-life examples and extensive lists of reference material the interested reader is referred to Albert and Barabási (2002).

Dorogovtsev and Mendes (2003) and Newman (2003). Statistical descriptions of financial networks are scarce in literature, however, which probably relates to the confidentiality of the transaction data. This may be especially true for payment systems. Exceptions are Soramäki, et al (2007), Lublóy (2006), Inaoka, et al (2004), Schmitz and Puhr (2007).

[13] The payment network is also a 'simple network'. It contains no self-loops (payer = receiver) nor parallel or multiple links from one sender to one other receiver. The links of the network form a set of node pairs (not just a collection, see Goodrich and Tamassia, 2006). For some applications this sense of direction is not essential. In that case, connections between nodes are formed by a single, undirected link (undirected network). This may for instance be the case when the establishment of a contact by a transaction is important, but not the direction of the transaction.

Figure 10.2 **Basic types of (undirected) networks**



10.4.2 Evolution of network properties

Payment systems are dynamic networks of which the number of nodes and links can vary greatly over time. The actual transfer of money only creates a temporary link. The choice of timescale for the statistical description of network properties is important and, for the Dutch case, the network properties of the dominant network component are representative for the whole payments network after about a 10 minute time period.[14] In this section we are interested in the properties of the network in 'normal times' and thus look one year of data from June 2005 to May 2006 (257 business days).

The network measures we analyse are explained in the Appendix.[15] They include network size in nodes and links, connectivity, reciprocity, path length, eccentricity, degree, degree correlation, degree distribution, nearest and second nearest neighbours, and clustering. The treatment of these properties aims at giving more insight in the topological structure of the network.

[14] After one (ten) minutes 36% (68%) of the data samples already consists of a single network component. In the case of the ten minute time frame, in the overwhelming majority of the 32% remaining cases there are only one or two minimally sized other components of two, three nodes.

[15] And in for example Dorogovtsev and Mendes (2003) and Soramäki, et al (2007).

Figure 10.3 displays the development over time of the various network measures *for the dominant network component*. For each of the measures, the x-axis represents the different time periods investigated (ie 1, 3, 5, 10, 30 minutes, 1, 3, 5 hours, 1, 3, 5, 15, 257 days) in minutes. The use of a logarithmic scale enables coverage of all time periods. It requires careful interpretation of the figures, though, to get a good understanding of the high rates of development for short time periods (≤ one hour) and lower rates for periods beyond one day. The discussion of the results is largely restricted to the relatively variable outcomes for the one hour time period and the relatively stable outcomes for the time period of one day. The former generally represent intraday network properties well, the latter network properties for periods from and beyond one day.

The figure shows major developments take place mostly in the first hour of network formation. From one hour to one day the network grows more gradually. The *size* of the network measured 88±6 nodes on an hourly basis and 129±5 nodes on a daily basis (top left). During the whole period of 257 days (only) 183 nodes have been active in total. These numbers characterize a small-size network, also with respect to other investigated banking networks.[16] In Inaoka, et al (2004) the total number of banks amounted to 354, in Boss, et al (2004) to about 900 and in Bech and Soramäki (2005) more than 5000 banks made up the system.

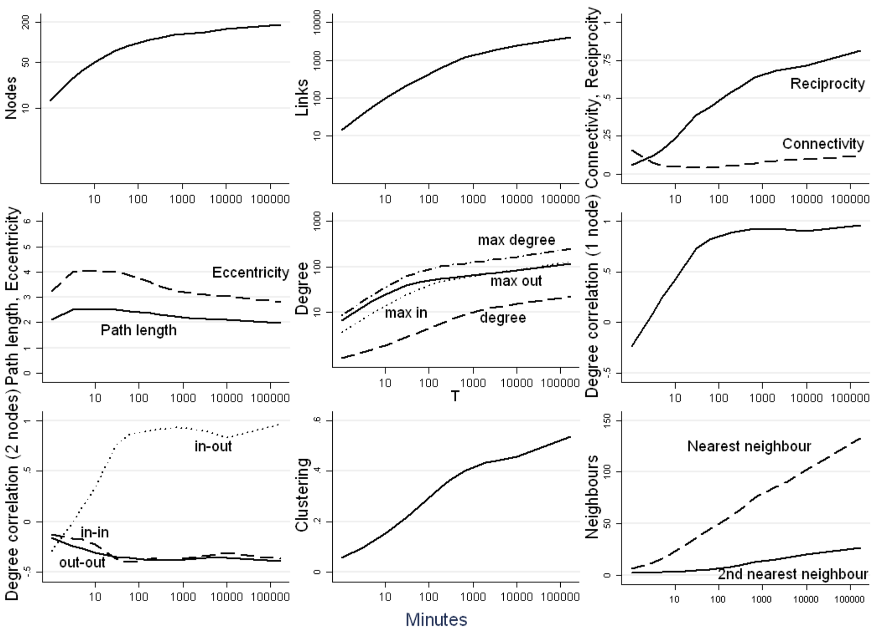
On an hourly basis 326±76 directed *links* were found between the nodes (top middle). On a daily basis there are almost four times as many links: 1182±61. The number of links increases at a higher rate than the number of nodes, but the number of *possible* links increases with the number of nodes *squared*. Over the whole period a mere 12% of the possible number of links between nodes (183*182=33306) actually did become a link, to a total of 4079 links for one or more transactions.

The fraction of actual to possible links, *connectivity*, gives better insight in the relative growth of nodes and links based on the proper, quadratic relation between them (top right). The values show that the network remains very sparse over all time periods. Connectivity rapidly declines from 0.16±0.12 after one minute to a minimum of 0.04±0.01 after approximately 30 to 60 minutes, to increase thereafter at a lower pace to 0.07±0.00 after one day and 0.12 after 257 days. The explosion of nodes in the first hour suppresses connectivity,

[16] Payments through TARGET to and from different banks in the same EU country are all recorded under the same country code and therefore belong to the same node. This leads to a downward bias.

because the growth of links does not keep up with the growth of nodes but after one hour the situation reverses. At all times, however, the network keeps its low connectivity and remains far from connected. Even after 257 days 88% of all theoretically possible links have not been used for a single transaction. *Reciprocity*, the fraction of links with a link in the opposite direction, displays a rapid increase in the first hour to 0.44±0.12 and increases at a lower rate to 0.63±0.02 after one day. It means that a link in one direction implies a high probability of a link in the opposite direction. Payments often take place in two directions. This, however, gives no information on the intensity of activity in both directions.

Figure 10.3 **Development of network properties over time (in minutes): nodes, links, connectivity reciprocity, path length, eccentricity, degree (also max degree, max in-degree, min degree, min in-degree) and degree correlation (for one node and for two nodes), clustering, nearest and second nearest neighbours**



The average *path length* between two randomly selected nodes forms another way to measure the size of the network (middle left). This distance peaks at 2.5±0.3 nodes at the beginning of network formation (5 min) to decline gradually thereafter to 2.0 nodes after 257 days. The latter is of the order of the logarithm of the size of the network (number of nodes), a feature predicted both for the classical equilibrium and fat-tailed, scale-free non-equilibrium networks (Dorogovtsev and Mendes, 2003). The numbers indicate that every node on average connects to another node through only one intermediate node. The maximum distance, *eccentricity* (or diameter), amounts to 4.1±0.7 after 5 minutes and declines gradually to 2.8. This means that over time the maximum number of steps participants have to take to reach the other participants decreases. Concentric, spherical connections gain strength in comparison to linear, radial connections. The network gets more structure and on a local level it becomes less 'tree-like'. The results again emphasize the small size of the network (in 'length' this time) and raise the question whether the intermediary node is also random in general, or that a core of central nodes exists through which other nodes connect.

Node *degree*, the number of links connected to a node, forms an essential measure for the description of the direct surroundings of a node (middle right). The degree measure can be split in in-degree and out-degree on the basis of the number of in- and outgoing links. The concept of a degree can easily be extended and generalized to concentric circles of neighbouring nodes with length 1, 2, .., n (n < network size). A close relationship therefore exists between degree and length of the network (Dorogovtsev and Mendes, 2003). Here the focus will only be on direct neighbours of length 1. From an initial value of 1 the network degree increases to 3.7±0.8 links per node after one hour and at a somewhat slower pace to 9.2±0.4 links per node after one day. It takes nearly the rest of the 256 days to (more than) double to 22.3 links per node.

These outcomes deviate significantly from many theoretical models of growing networks which assume a fixed degree (linear growth). In these models each added node is accompanied by a fixed number of new links (eg Barabási and Albert, 1999, see also discussion in Dorogovtsev and Mendes, 2003). The payment network clearly exhibits a form of accelerated growth, because the degree increases during network growth. The model of network growth would also differ from theoretical models due to the upper boundary in the number of participants. Growth in nodes inevitably declines over time, since fewer nodes can be added. The theoretical model of the payment network, including accelerated growth in links and a declining growth in nodes due to the limited number of participants, is a subject for further study.

The time development of the *maximum degree* shows node degree covers a large range of values across the network (middle right). The maximum degree increases from a level of about 9 times the average degree after 1 minute, to a level of around 20 for periods between ten minutes and three hours, slowly declining over time to a level of 11 afterwards. Concretely, it means that after one hour the average node may hold 3.7±0.8 links, but the maximally linked node actually holds 79±13 links. *Maximum out-degree* surpasses *maximum in-degree* for periods up to a day. These maxima reflect the presence of one or more highly connected nodes. The presence of large differences in degree values also reflects large differences in the local network structure, hinting at a structure of many low-degree and some high-degree nodes. The actual degree distribution of nodes across the network, discussed at the end of this section, therefore contains indispensable information about the local structure.

The *degree correlation* (centre) between in-degree and out-degree of individual nodes starts off negatively but becomes very strongly positive after just thirty minutes ([70%–100%]). This means that above (below) average in-degree has a high chance of being accompanied by above (below) average out-degree. Nodes that make payments to many counterparties also receive payments from many counterparties. The results on reciprocity already showed nodes are often counterparties in both directions. Degree correlations between in-degree and out-degree of two connected nodes largely follow the same pattern (bottom left). Degree correlations between in-degree respectively out-degree of two connected modes prove negative. The results on degree correlations again suggest the existence of a few strongly connected nodes linking to several weakly connected nodes.

The *clustering coefficient* (bottom middle) measures the probability of two neighbours of a node sharing a link among themselves, too. Where distance measures length, clustering measures density of the network structure at a local level. It gives information about the direct surroundings of the nodes. As expected, the development of the clustering coefficient over time confirms that formation of connections across neighbours takes more time to develop than sheer growth of the network. Still, the rate of increase in clustering is relatively high in the first hour and somewhat lower afterwards. The average clustering coefficient increases from 0.26±0.09 after one hour to 0.40±0.02 after one day. After 257 days, average clustering amounts to 0.53. It means that, on average, in half of the cases the neighbours of a randomly chosen node are connected among themselves, too. When comparing this to classical equilibrium networks, the numbers indicate relatively high correlations in the form of clustering exist on a local level. As in Soramäki, et al (2007) the number of nodes with a clustering coefficient 0, however, is very high: 49% after one hour, 27% after one day. These can mostly be attributed to nodes with only one or just a few links, because the probability of at least a single link across neighbours increases rapidly with the number of links (more precise: the number of combinations of neighbours increases with the number of links *squared*). In any case, local density proves absent for a significant part of the network on a local level.

The last pane (bottom right) displays the number of *nearest neighbours* z_1 and second *nearest neighbours* z_2 of a node. In determining z_1 and z_2 the direction of the links has been ignored,[17] since it is the presence of a connection that matters here (not the direction of it). The number of nearest and second nearest neighbours amounts to 5.7±0.1, respectively 45±4 after one hour. These numbers increase to 12.5±0.5, respectively 75±3 after one day. They emphasize that the number of direct contacts z_1 strongly increases over time and that the second line of contacts z_2 through z_1 (logically) is a multiple of the first line (the number of contacts grows strongly between the first and second line). Like in the case of the clustering coefficient, the results confirm that local structure takes more time to develop than the size of the network.

We now turn to *degree distributions*. As mentioned earlier, these distributions give indispensable information about the relative 'popularity' of participants in the system. Large banks are obvious examples of popular (highly connected) participants, but also the specific clearing institution which settles many, relatively small, customer driven payments. As such payments go to and from most institutions, this clearing institution will be a highly connected node many participants attach to.

The concept of degree can easily be extended from nearest neighbours to second nearest neighbours, to third nearest neighbours, etc, but here the focus will be on nearest neighbours only. In Figure 10.4 the degree is plotted on the x-axis with the associated probability on the y-axis (both on a logarithmic scale).[18] The darker the dot associated with each degree bucket, the larger the size of the firm(s) in that bucket. We measure firm size by total annual transaction value.

For one hour time snapshots the distribution already steeply declines at very low degree values; for one day time snapshots this

[17] This distinguishes z_1 from the degree.
[18] Averaging has taken place over all snapshots in the dataset.

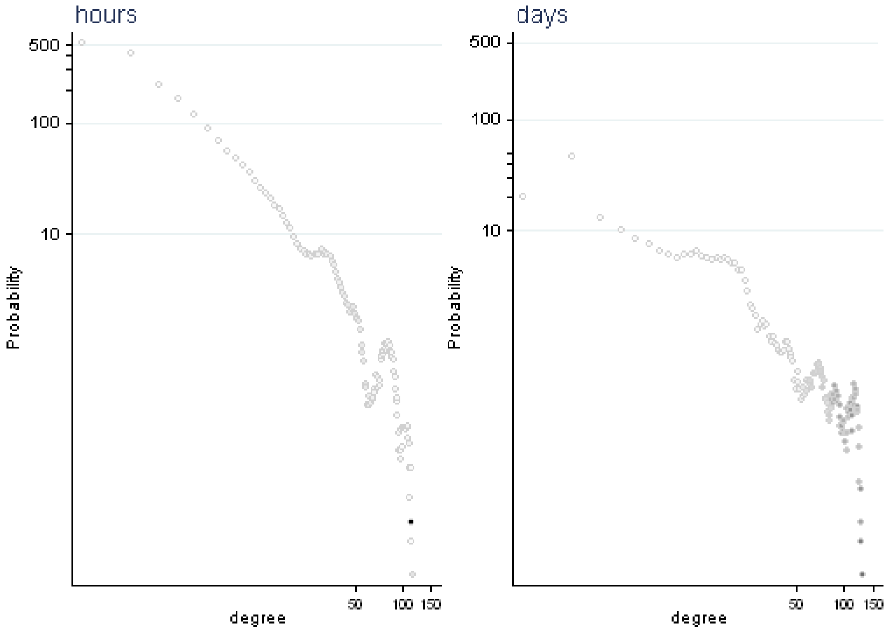
rapid decline starts off from a degree value of about twenty. The largest observed size of a one hour (day) network amounted to 116 (130) nodes. The networks are characterized by a high number of nodes with relatively few connections and a small number of (relatively) highly connected nodes. Each of the smaller humps at the high degree end of the x-axis accounts for an individual node or a small group of individual nodes. These humps are basically distributions of individual (groups of) nodes.

Further, in going from one hour to one day time snapshots the frequency of highly connected nodes increases at the expense of weakly connected nodes. During the treatment of the clustering coefficient and degree correlations it was already mentioned that connections across neighbours take more time to develop than growth of the network. The local structure becomes stronger with time. Also, the degree distributions for the payment network cover too few order of degree (≈ 2), to perform any fitting to a power-law distribution (to test whether the distribution is a scale-free, fat-tailed distribution).[19]

The shading of the dots in the graph tells us that for shorter time periods, in this case the one hour snapshots in the left pane, the firms with the highest degrees are not necessarily the largest firms (in terms of value transferred). The left pane clearly shows the importance of participants that handle batches of consumer payments; the value of these payments are not high but they do entail many connections. In comparison, the one day snapshots in the right pane show us that high degrees are associated with large turnover.

[19] See §5.6 and footnote 11 of Dorogovtsev and Mendes (2003) for a critical note on the empirical 'observations' of power-law distributions. Scale-free networks have a size-dependent cut-off, which sets strong restrictions for such observations over 2 or 3 orders of degree.

Figure 10.4 **Degree distributions for time snapshots of one hour and one day, respectively**



10.5 Vulnerability of the network structure

In the introduction we noted that the study of the topological structure of the payment network (or any other network) is not a goal in itself, but a means for understanding the processes that make use of the structure. A particularly interesting topic of research is the vulnerability (or resilience) of the network to random or directed failures. The impact of a failure of a single node may remain confined locally or cause a shockwave that propagates through the system (systemic risk). The purpose here is to show that network theory provides tools for studying this risk.

We analyse the impact on network properties of removal, one by one, of the most highly connected nodes (cf Albert, et al, 1999).[20] Risks to the system may surface upon showing the importance of specific nodes to the topological structure. The removal procedure is equivalent to building the network from the raw transaction data, but

[20] See Heijmans (2009) for a different sensitivity test involving a simulated response to stress situations.

leaving out all individual transactions that involve the specific 'removed' nodes. It is a static procedure with shortcomings, like for instance the absence of any adaptive behaviour. The topology after removing n specific nodes is always the same, but the path in getting there will differ upon changing the order of removal of the nodes. It means the procedure will not identify a unique dependence of the topology on any of the individual, removed nodes. What the procedure does bring to light is the dependence of the topology on removal of a limited number of highly connected nodes.

Following the literature we choose one day as the timescale to measure the network properties. Figure 10.5 shows the impact of the removal procedure on several selected network properties. The x-axes show network properties in the initial situation ('0') and, in going to the right, those properties after removing the most highly connected node ('–1'), the second most highly connected node ('–2'), etc, until ten nodes have been removed ('–10'). The initial situation ('0') for all properties is the same as in Figure 10.5 (one day time period).

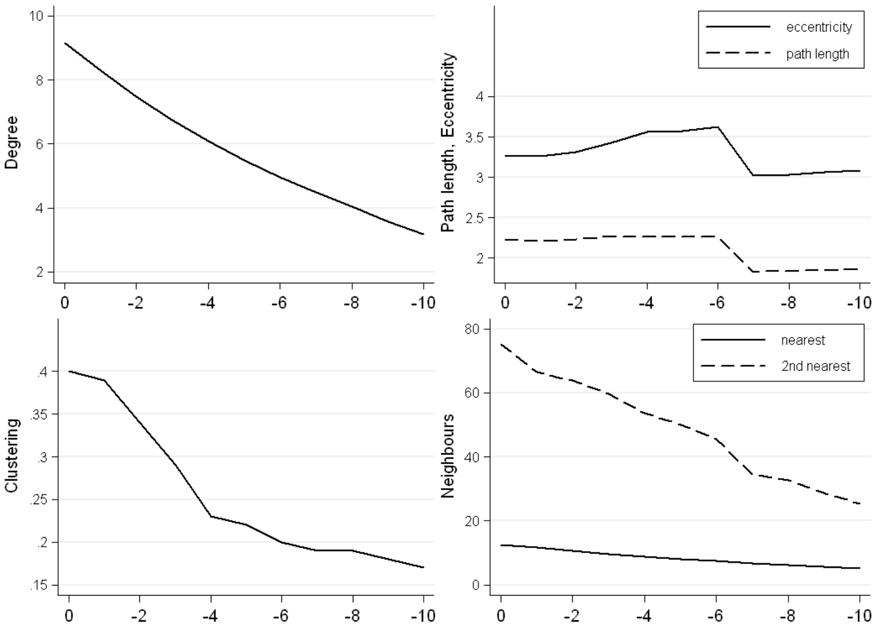
The network becomes smaller and even more sparse as for instance shown by the *degree* values (top left corner) which decrease steadily, except for point '–7', from 9.2 to 4.3. This results from the number of nodes having declined from 129 to 89 (–31%) and the number of links from 1182 to 378 (–68%). Moreover, connectivity decreases from 0,072 to 0,049. The network loses more nodes than the 10 deliberately removed nodes, because on average 30 neighbouring nodes with a single link will lose their last connection during this procedure. The seventh node ('–7') is a good example, since the degree actually increases upon removing this node.

The removal of central, highly connected nodes increases the path lengths between the remaining nodes. In the removal of the seventh node this phenomenon is outweighed by the accompanying loss of the single link nodes and the shortest paths between them and all other nodes. Specifically, *path length* and maximum path length (top right corner), or *eccentricity*, increase from 2.2 to 2.5 and from 3.3 to 4.2, respectively.

The outcomes for clustering and correlations both show the local structure starts to break down (bottom left corner). Clustering, or density of connections on a local scale, decreases from 0.40 to 0.23. The removal of nodes two to four has an unevenly negative impact on clustering in comparison to the other nodes. The out-out degree correlation increases more steadily from –0.38 to –0.14 (= loss of correlation).[21] The outcomes for nearest neighbours and second nearest neighbours confirm this breakdown in structure (bottom right corner).

The impact of removing the ten most highly connected nodes on the key aggregate figures of the payment network is severe. Value transferred and number of transactions decline steeply to respectively only 6% and 12% of the initial situation (not shown here). This marks these nodes as essential to the core function of the payment network. This holds especially for the first 4 nodes, since by their removal value transferred and the number of transactions have already declined to 27% and 30%, respectively, of the initial situation.

Figure 10.5 **Impact of node removal on network properties: degree, path length, eccentricity, clustering, out-out degree correlation, nearest and second nearest neighbours (z_1 and z_2)**



[21] In-in degree correlation increases from –0.38 to –0.10. In-out degree correlation decreases from 0.93 to 0.59.

It should be clear that random removal of ten nodes would not have caused the same impact on the network structure and key aggregate figures. In this sense the results are comparable to those in Albert, et al (1999) in that the system is vulnerable to a directed failure (here: removal of a highly connected node) due to the importance of the relatively highly connected nodes in the tail of the degree distribution. In addition, the discussed procedure of node removal convincingly shows network theory provides tools for analysing distortions to the network.

10.6 The network structure during recent turmoil in credit markets

In the previous section the vulnerability of the network has been illustrated principally by a static, hypothetical procedure of node removal. The dependence of network properties on the most highly connected nodes proved to be strong. Other, more realistic events can also affect the proper functioning of the payment system. A prime example is a possible loss in confidence between banks which would reduce the liquidity of funds in the markets as we have seen in the recent crisis. If banks delay or stop making payments to (some) other banks, this will have its effect on the functioning of the payment system if the scale of such change in behaviour or the scale of banks involved is large enough.

Given the magnitude of the crisis, the impact on the TARGET2 system as a whole seems to have been limited. Turnover increased with 13% reflecting, first, the increased provisioning of liquidity and, second, the shift to shorter maturity loans (ECB 2008). Heijmans et al (2011) do however not find a shift from longer to shorter term lending in the Dutch unsecured interbank money market. Looking at the pattern of payments throughout the day, value and timing of payments are relatively unchanged. Increased turnover is concentrated in the last hour of operations, mainly for overnight deposits at central banks. In contrast with other infrastructures, it was not necessary to change operating hours to allow banks to settle back logs.

Does the limited impact on the European level, as for instance reflected in the traditional measures like turnover, also carry over to network measures? Or, put differently, do these measures provide us additional insights? To answer these questions, we compare the network statistics for the apex of the crisis (2008) with those of earlier emerging stress (2007).[22] A complication in this comparison of in- and pre-crisis data is the implementation of the TARGET2 system. This platform has been phased in as of 19 November 2007 with the Netherlands migrating on 18 February 2008. As a result, payments and hence network structure have changed. In the old TARGET, internationally active banks would allocate liquidity in the morning only to concentrate it in the evening for the central treasury. In addition, many such banks now no longer allow subsidiaries to settle local payments in local settlement systems but rather use a single TARGET2 account. Thus, many transactions would no longer be recorded locally and therefore the entry point in to the TARGET2 system does not necessarily correspond very closely to the geographical location of the account holder.

To make the comparison as accurate as possible we compare the 18 February 2008 – 31 December 2008 period with the same period the year before. The 2008 crisis thus only contains TARGET2 payments. Figure 10.6 displays the difference (2008 value – 2007 value) for each of the measures shown earlier in Figure 10.4. The graphs show that the network structure has changed significantly: participation, as measured by the number of nodes has increased almost fivefold (after 1 day). The number of links is up a little less than fourfold. This does not reflect a sudden inflow of banks using the system; rather, foreign banks now have their own account number instead of being subsumed in the older country nodes. The links formed have become a little more dispersed as shown by the drop in connectivity.

Drawing strong conclusions based on the Figure 10.6 might be dangerous however as the changes shown might merely reflect changes in the way payments are recorded induced by the migration to TARGET2 rather than banks changing their behaviour in response to the crisis. To investigate this issue it is helpful to look at the development of the network measures over time. In principle we could analyse each measure at each time interval but for computational reasons we limit ourselves to the 1 day interval. Incidentally this is the interval commonly used in the literature. Overall the analysis of the graphs show that there is significant variability in network measures over time. On many occasions the changes are related to entry of new participants or other technical changes. Although the measures thus reflect actual changes in the network, they are less suitable for

[22] Admittedly the latter part of 2007 could also be characterised as a crisis period.

determining the effect of the crisis. Ignoring the more elaborate network measures for the moment and looking at the development over time of the more traditional measure Gross Turnover, we see that the demise of Lehman (15 September 2008, indicated by the red line in Figure 10.7 does seem to have triggered a period of heightened activity. More importantly, the figure also shows that this activity was not equally spread-out over the individual banks. It is therefore still important to monitor the individual firms.

Figure 10.6 **Changes in traditional system measures and network properties over time (2008 minus 2007 values)**

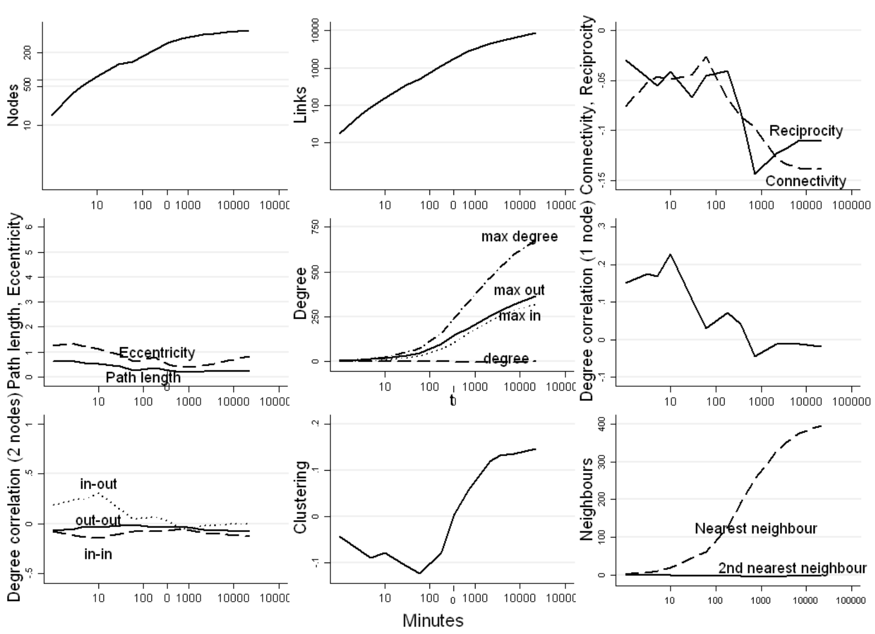
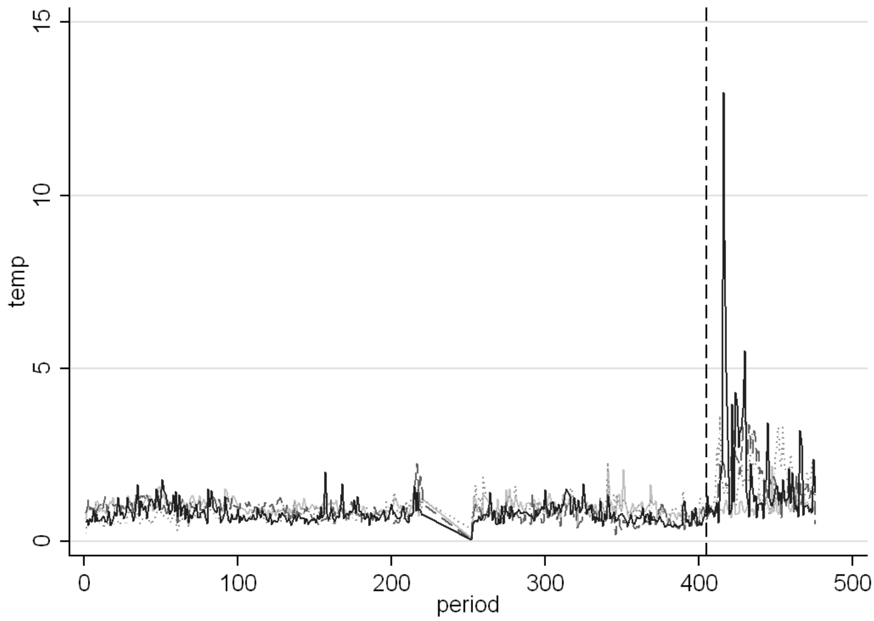


Figure 10.7 **Development of a selection of traditional system measures and network properties over time for the largest banks (2007–2008): transaction value and degree**



Note: The data period runs from 1 January 2007 to 31 December 2008. The first two months in January and February 2008 are currently not show but will be included in the future.

10.7 Conclusions and the way forward

Recently, interest in the topological structure of networks has risen significantly. Application of the ideas of network theory to payment systems is still limited, though. This study adds to literature in showing the measures available to characterize such networks. The application of these measures illustrates the influence of the chosen time frame on the properties of the payment network and the central role of highly connected banks in the functioning of the payment network. Moreover, the migration to TARGET2, with its associated changes in payment behaviour, has resulted in a dramatic change of the measured network.

We first gave a description of the old TOP-network in terms of 'traditional' measures such as turnover and average payment size. An international comparison with other payment systems revealed that the analyzed network is midsized and relatively active. Looking at intraday developments activity proves high right after opening, mainly due to queued orders entered the previous day. With a brief rise in activity right before lunch, the all day high in value transferred takes place between four and five o'clock.

We proceeded with a presentation of the time development of several important network measures. Time is crucial in analysing networks with short lived links, like payment networks. The outcomes have made explicit to what extent fast development takes place in the early phase of network formation of about one hour and slower development afterwards. The payment network proves to be small (in nodes and links), compact (in path length and eccentricity) and sparse (in connectivity) for all time periods. Measurements of degree and degree correlations, clustering, and the number of nearest and second nearest neighbours describe the development of network structure at a local level. As expected, development of network structure takes more time than growth of the network in terms of size.

The actual degree distribution contains indispensable information on the local structure. It proves that the network is characterized by a high number of nodes with relatively few connections and a low number of highly connected nodes. Of particular importance is the observation that the average degree increases during growth of the network. This contrasts with many theoretical models that assume node degree remains fixed. Further work could therefore be directed at modelling the payment network using a type of accelerated network growth in links.

We showed that the payment system is vulnerable to a directed failure. The vulnerability of the network was tested by removing, one by one, the ten most highly connected nodes. Node removal had a strong impact on value transferred, number of transactions and network properties like degree, path length and eccentricity, clustering and degree correlation, and nearest and second nearest neighbours. These outcomes emphasized the central role the most highly connected banks, especially the top 4 of these, play in the payment system: they are essential to the core function of the payment network.

We also investigated whether the start of the financial crisis has led to changes in the structure of the payment network by analysing time-series of network measures. We concluded that network structure of the payment system has changed materially following the migration to TARGET2. It is therefore difficult to determine if and how the crisis seems not to have affected the network. Since (severe) disruptions in the payment system would inevitably show up in the discussed measures, it is nevertheless useful to monitor for changes in traditional system measures and in network properties. A preliminary analysis of time-dynamics seems to indicate that network characteristics have changed following the demise of Lehman.

The current study intends to show how various measures can be used in analysing payment networks. It is also an exploratory study: two clear directions for further research are analysing, first, the importance of link weights and, second, the role of collateral and available liquidity in absorbing shocks.

On a more conceptual level this study also highlights a significant gap in our understanding. This analysis has very little to say about how changes in behaviour (driven by eg changes in risk preferences) would lead to changes in the network structure. Especially under stress (ie (potential) network breakdown), it is likely that behaviour will change radically. Incorporating behaviour in our models remains a challenge. Network simulations, based on the empirical network structure, might be a possible way forward. Ultimately, knowledge of the functioning of the payment network is crucial for preserving its stability. On a practical level our study shows that demands on the quality, quantity and timeliness of data are significant.

Appendix

Network properties[23]

Size

The most basic network properties are the number of nodes *nodes* (n) and *links* (l). The former is often referred to as the size of the system. The relative number of links l to the possible number of links determines the network *connectivity* (c). It represents the probability of two nodes sharing a link. For a directed network, with links between nodes in two directions, connectivity is given by $c = l/(n \cdot (n - .1))$. For a connected network (ie without disconnected components) $l \geq n - 1$. In the special circumstance $l = n - 1$ the network is a so called tree network with minimal connectivity $c = 1/n$. Connectivity reaches it maximum value $c = 0$ for a completely connected network. All possible links have then been realized. Reciprocity, finally, is the fraction of links with a link in the opposite direction (range from 0 to 1).

Path length

A *path* is an alternating sequence of connected nodes and links that starts and terminates at a node. If all links represent unit length, *path length* l_{ij} between nodes i and j is the length of the shortest path between the nodes. The average path length l_i for node i is the average distance to all other nodes. Although a directed network in principle consists of directed paths that are being traversed in the direction of the links, direction is not taken into account here. The path represents a connected sequence of contacts in the form of transactions rather than a sequence of directed flows of payments. Link weights in terms of value transferred may vary strongly over one path so that direction of flow, without explicitly taking into account link weights, not necessarily contains very valuable information. *Average network length* l_{avg} is the average of all path lengths l_i. It determines the average undirected shortest path. Network *eccentricity* (e) is defined as the largest of the observed path lengths: $e = \max_i, j(l_{i,j})$.

[23] Based on Dorogovtsev and Mendes (2003) and Soramäki, et al. (2007).

Degree

The number of links between one node i and other nodes determines the *node degree* (k_i). In a directed network these connections consist of incoming and outgoing links, which respectively determine the *in-degree* ($k_{in,i}$), the *out-degree* ($k_{out,i}$), and *node degree* (k_i) by $k_i = k_{in,i} + k_{out,}i$. Every link contributes exactly one unit to both the out-degree of the node at which it originates and to the in-degree of the node at which it terminates. The *average degree* (k_{avg}) of a network is the relative number of all links to all nodes: $k_{avg} = 1/n = 1/2n\sum_i k_i = 1/n\sum_i k_{in,i} = 1/n\sum_i k_{out,i}$.

The *maximum in-degree* $k_{in,max} = \max_i(k_{in,i})$, *maximum out-degree* $k_{out,max} = \max_i(k_{out,i})$ and maximum degree $k_{max} = \max_i(k_i) = \max_i(k_{in,i} + k_{out,i})$ determine the maximum degree values and the maximum deviations (to the upside) from the respective average degree values. More informative and more elaborate to determine are the degree distributions $P(k_i)$, $P(k_{in,i})$ and $P(k_{out,i})$ for a specific node i. Summation over all nodes i and taking averages results in the total degree distributions P(k), $P(k_{in})$ and $P(k_{out})$. Two examples of degree distributions are respectively a Poisson distribution and a power-law distribution. The former results when a fixed number of nodes is randomly connected on the basis of the fixed network degree k_{avg}. Larger networks asymptotically follow the Poisson distribution $P(k) = \frac{e^{-k_{avg}} \cdot k_{avg}^k}{k!}$ (classical equilibrium network). In practice, however, many networks have (relatively recently) been found to follow a power-law distribution $P(k) \propto k^{-\gamma}$. These non-equilibrium networks are characterized by fat-tails which mark the relatively high frequency of highly connected nodes in comparison to classical equilibrium networks. They originate from growing networks in which new nodes (linearly) preferentially attach to other nodes. They have no natural scale and are called scale-free networks. In recent years it has been demonstrated that many social, informational, technological and biological networks have fat-tailed, scale-free degree distributions (see for instance Amaral, et al, 2000, Newman, 2003 and Dorogovtsev and Mendes, 2003).

Degree correlations

Degree correlations between neighbouring nodes provide additional information on the network structure. In an uncorrelated network the degree of one node is independent of its neighbouring nodes. Degree correlations therefore provide information on whether nodes are generally connected to nodes with comparable degree, to nodes of different degree, or if there is no relation at all. Classical random networks have no correlations. Fat-tailed, scale-free networks on the other hand may exhibit strong correlations.

Several measures exist for degree correlations. For example

– Between k_{in} and k_{out} for individual nodes
– Between k_{in} and k_{out}, k_{in} and k_{in}, or k_{out} and k_{out} for two nodes

Clustering coefficient

Another concept to describe the correlation between nodes is the *clustering coefficient* (C_i), which gives the probability that two neighbours of a node share an undirected link among themselves. It marks the density of connections in the direct neighbourhood of a node (cliquishness). The clustering coefficient is determined by the number of actual undirected links between nearest neighbours ($l_{nn,i}$) of a node i as a fraction of the number of possible undirected links: $C_i = \frac{2l_{nn,i}}{k_i \cdot (k_i - 1)}$. The average clustering coefficient (C_{avg}) over all nodes determines the network clustering. The meaning of the coefficient becomes particularly clear in a social context where it is the extent of the mutual acquaintance of friends. The clustering coefficient ranges from 0 for a tree network to 1 for a completely connected network. The classical random network locally has a tree-like structure (loops cease to exist in the infinite network). Fat-tailed, scale-free networks may exhibit strong clustering.

References

Albert, R (1999) **Diameter of the World Wide Web.** Nature, 401, 130–131.

Albert, R – Barabási, A-L (2002) **Statistical Mechanics of Complex Networks.** Reviews of Modern Physics, 74, 47–97.

Albert, R – Jeong, H – Barabási, A-L (1999) **Error and Attack Tolerance of Complex Networks.** Nature, 406, 378–382.

Allen, F – Gale, D (2000) **Financial Contagion.** Journal of Political Economy, 108, 1–33.

Amaral, L A N – Scala, A – Barthélémy, M – Stanley, H E (2000) **Classes of Behaviour of Small-World Networks.** Proceedings of the National Academy of Sciences of the United States of America, 97, 11149–11152.

Barabási, A–L – Albert, R (1999) **Emergence of Scaling in Random Networks.** Science, 286, 509–512.

Bech, M L – Garrat, R (2006) **Illiquidity in the Interbank Payment System Following Wide-Scale Disruptions.** Federal Reserve Bank of New York Staff Reports, 239.

Bech, M L – Soramäki, K (2005) **Systemic Risk in a Netting System Revisited.** In Liquidity, Risks and Speed in Payment and Settlement Systems – a Simulation Approach, ed. by Leinonen, H: Bank of Finland Studies, 151–178.

BIS (2007) **Redbook: Statistics on Payment and Settlement Systems in Selected Countries.**

Boss, M – Elsinger, H – Summer, M – Thurner, S (2004) **The Network Topology of the Interbank Market.** Quantitative Finance, 4, 677–684.

Dorogovtsev, S N – Mendes, J F F (2003) **Evolution of Networks.** Oxford: Oxford University Press.

ECB (2008) **TARGET Annual report.**

Erdõs, P – Rényi, A (1959) **On Random Graphs.** Publicationes Mathematicae, 6, 290–297.

Faloutsos, M – Faloutsos, P – Faloutsos, C (1999) **On Power-Law Relationships in Internet Topology.** Computer Communications Review, 29, 251–262.

Furfine, C H (1999) **The Microstructure of the Federal Funds Market.** Financial Markets, Institutions and Instruments, 8, 24–44.

Goodrich, M T – Tamassia, R (2006) **Data Structures and Algorithms in Java.** London: John Wiley & Sons.

Goyal, S – Leij, M V D – Moraga-gonzález, J L (2006) **Economics: An Emerging Small World.** Journal of Political Economy, 114, 403–412.

Heijmans, R (2009) **Simulations in the Dutch Interbank Payment System: A Sensitivity Analysis.** DNB Working Paper Series, 199.

Heijmans, R – Heuver, R – Walraven, D (2010) **Monitoring The Unsecured Interbank Money Market using TARGET2 data.** DNB Working Paper, 276

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 Paper, 04-E-04.

Ledrut, E (2006) **A Tale of the Water-Supplying Plumber: Intraday Liquidity Provision in Payment Systems.** DNB Working Paper Series, 99.

Lublóy, Á (2006) **Topology of the Hungarian Large-Value Transfer System.** Magyar Nemzeti Bank Occasional paper, 57.

Newman, M E J (2003) **The Structure and Function of Complex Networks.** SIAM Review, 45, 167–256.

Oord, A V – Lin, H (2005) **Modelling Inter- and Intraday Payment Flows.** DNB Working Paper Series, 74.

Pericoli, M – Sbracia, M (2003) **A Primer on Financial Contagion.** Journal of Economic Surveys, 17, 571–608.

Rosati, S – Secola, S (2005) **Explaining Cross-Border Large-Value Payment Flows – Evidence from TARGET and Euro I Data.** ECB Working Paper Series, 443.

Soramäki, K – Bech, M L – Arnold, J – Glass, R J – Beyeler, W E (2007) **The Topology of Interbank Payment Flows.** Physica A, 379, 317–333.

Schmitz, S – Puhr, K (2007) **Structure and Stability in Payment Networks – a Panel Data Analysis of ARTIS Simulations.** Working Paper.

Van Lelyveld, I P P – Liedorp, F R (2006) **Interbank Contagion in the Dutch Banking Sector.** International Journal of Central Banking, 2, 99–134.

Chapter 11

Systemic risk in large value payment systems in Colombia: a network topology and payments simulation approach*,**

*Carlos León*** – Clara Machado**** − Freddy Cepeda***** − Miguel Sarmiento*******

* An earlier (longer) version of this paper was selected to be presented at the Second BIS Consultative Council for the Americas Conference, Ottawa, Canada, May 2011. Authors are grateful to the Board of Directors of Banco de la República, Gerardo Hernández, Hernando Vargas, José Tolosa and Dairo Estrada for their comments and suggestions to this or to a preliminary document. Technical assistance and support by Joaquín Bernal was decisive for the development of the project and this document. Work by Orlando Chipatecua and Jorge Cely in a preliminary document is also acknowledged and appreciated. Additional research assistance provided by Gabriel Angarita. As usual, the opinions and statements are the sole responsibility of the authors.

** Results are intended for academic and research purposes; they may not be used to infer credit quality or any type of assessment for any financial institution.

*** Research and Development Section Manager, Financial Infrastructure Oversight Department, Payments Systems and Banking Operation Division (PSBOD), Banco de la República (BR), cleonrin@banrep.gov.co/ carlosleonr@hotmail.com (corresponding author).

**** Head of the Financial Infrastructure Oversight Department, PSBOD–BR, cmachafr@banrep.gov.co.

***** Leader Professional, Financial Infrastructure Oversight Department, PSBOD–BR, fcepedlo@banrep.gov.co.

****** Expert Professional, Operations and Market Development Department, BR, nsarmipa@banrep.gov.co.

11 Systemic risk in large value payments systems in Colombia: a network topology and payments simulation approach

Abstract

The most recent episode of market turmoil exposed the limitations resulting from the traditional focus on too-big-to-fail institutions within an increasingly systemic-crisis-prone financial system, and encouraged the appearance of the too-connected-to-fail (TCTF) concept. The TCTF concept conveniently broadens the base of potential destabilizing institutions beyond the traditional banking-focused approach to systemic risk, but requires methodologies capable of coping with complex, cross-dependent, context-dependent and non-linear systems.

After comprehensively introducing the rise of the TCTF concept, this paper presents a robust, parsimonious and powerful approach to identifying and assessing systemic risk within payments systems, and proposes some analytical routes for assessing financial authorities' challenges. Banco de la Republica's approach is based on a convenient mixture of network topology basics for identifying central institutions, and payments systems simulation techniques for quantifying the potential consequences of central institutions failing within Colombian large-value payments systems.

Unlike econometrics or network topology alone, results consist of a rich set of quantitative outcomes that capture the complexity, cross-dependency, context-dependency and non-linearity of payments systems, but conveniently disaggregated and dollar-denominated. These outcomes and the proposed analysis provide practical information for enhanced policy and decision-making, where the ability to measure each institution's contribution to systemic risk may assist financial authorities in their task to achieve payments system's stability.

11.1 Introduction

The most recent episode of market turmoil exposed the limitations resulting from the traditional focus on too-big-to-fail (TBTF) institutions within an increasingly systemic-crisis-prone financial system. It is clear now that financial stability may not only be endangered by massive banking institutions, but also by non-banking participants significantly and intricately linked within the payments system. This has encouraged the appearance of the too-connected-to-fail concept (TCTF), and has fostered an ongoing debate on financial authorities' (ie central banks, regulators and supervisors) role facing systemic shocks, either during market's disruption or tranquil periods.

Acknowledging the TCTF concept broadens the base of potential destabilizing entities beyond the traditional banking-focused approach to systemic risk, but requires methodologies which cope with complex, cross-dependent, context-dependent and non-linear systems. A current trend for assessing the complexity and cross-dependency of financial and payments systems is based on network topology (NT), whilst context-dependency and non-linearity tends to be overlooked.

Hence, despite providing a comprehensive picture of systems' stability and resilience, NT is not suitable for approaching some of financial authorities' key practical concerns: If a systemic relevant institution fails, what is the intra-day and end-of-the-day dollar-value of the liquidity required by each institution within the system? Is the legal framework for customary and last-resort liquidity facilities appropriate for all the system´s participants? Is there any single institution or type of institution that conceals systemic risk? What is the market's liquidity level which may intensify dependence between institutions?

Therefore, based on a convenient mixture of NT (Becher et al, 2008, Soramäki et al, 2006) and payments systems simulation techniques (Leinonen and Soramäki, 2004), Banco de la República (BR) developed a robust, parsimonious and powerful approach for identifying and assessing systemic risk within Colombia's financial markets. First, NT basics are used to identify TCTF institutions according to the centrality concept. Afterwards, based on the observed transactions of an estimated payments system's typical day, the simulation procedure replicates Colombian large-value payments systems' queue resolution and multilateral settlement algorithms in order to quantify the potential consequences of the collapse of a TCTF institution.

Unlike econometrics, other customary approaches and NT alone, results consist of a remarkably rich set of quantitative outcomes which capture the complexity, cross-dependency, context-dependency and non-linearity of the payments system, but conveniently disaggregated and dollar-denominated. These outcomes and the proposed analysis provide financial authorities with practical information for enhanced policy and decision-making, where the ability to estimate each institution's contribution to systemic risk may assist financial authorities in their task to achieve payments system's stability.

This paper is divided in tree sections. The first – next-section summarizes briefly main aspects in the evolution for detecting and assessing systemic risk from the TBTF concept to the recent appeal TCTF concept, whose rationale behind the surge of connectedness. The second section is dedicated to familiarize the reader with some key features of Colombia's payments system; to present the chosen approach, and to analyze the results. Finally, the third section makes some concluding remarks that may be useful for financial authorities.

11.2 From too-big-to-fail concept to too-connected-to-fail concept for systemic risk

Traditional assessment of systemic risk has focused on those market participants considered as TBTF, where that label may be granted to an institution when, due to its size, its inability to meet its obligations could result in the inability of other system participants or of financial institutions in other parts of the financial system to meet their obligations as they become due. Basically TBTF institutions are those exceeding an asset-size cutoff (Saunders et al, 2009), which is a convenient and straightforward metric readily available for any regulator or central bank, even accessible for any market participant or a fairly informed ordinary man.

Despite more complicated definitions may focus on the volume of financial services (eg deposits, loans) provided by an institution within the financial system (IMF et al, 2009) or other less forthright metrics, the TBTF concept for identifying systemically important institutions is rather uncomplicated, and may explain why customary tools for crisis prevention and management are designed specifically for large bank runs (eg lender of last resort – LLR –, deposit insurance). Moreover, because it focuses on standard accounting data (eg assets, investments, deposits), financial authorities have found this approach as practical. This type of supervision may be depicted as micro-prudential, since, as defined by Brunnermeier et al (2009), it focuses on factors that affect the stability of individual institutions.

After the crisis literature has converged to declare the obsolescence of the current model of supervision and systemic risk assessment. Banks, which have been considered as the main focus of systemic risk detection and assessment because of their size, were not the main source of systemic risk as before (eg via non-performing loans, mismanagement of assets, balance mismatch). The financial system has changed dramatically since the Great Depression: though banks still play a large role, many functions that defined their traditional domain are increasingly performed by securities markets and non-bank market participants (Kambhu et al, 2007), namely unleveraged institutional investors (eg mutual and pension funds) and highly leveraged institutions (eg hedge funds); this is the 'shadow banking system' mentioned by Krugman (2009) and Acharya et al (2009).

Thus, evidence demonstrates that focusing on the institutions' size averts authorities from effectively detecting and assessing the systemic risk lurking beneath the nowadays highly complex and interconnected global financial system; this is, regulation and supervision were too institution-centric to see through to the systemic risk (IMF, 2009b). Hence, several authors (Chan-Lau, 2010, Clark, 2010, Acharya et al, 2009, Saunders et al, 2009, Zhou, 2009, Brunnermeier et al, 2009, Trichet, 2009) recognize the inevitability of using a broader set of concepts in order to detect and assess systemic risk.

Finally, it is possible to conveniently characterize financial system's issues and challenges as follows: the sum of complexity, homogeneity and opaqueness results in a robust-yet-fragile-and-uncertain system, where the existence of a defective risk management framework and the absence of liquidity facilities able to cope with the shift towards market liquidity risk make the financial system highly prone to systemic crisis.

Models oriented to detecting and assessing systemic risk within the financial system's complexity and homogeneity framework previously described are rather new. Current regulation is focused not on systemic risk, but rather on the individual institution's risk (ie micro-prudential), whereas regulation encourages financial institutions to distribute their risks in an unfettered manner around the system and to unregulated entities, which leads to excessive systemic risk (Acharya et al, 2009).

The most recent global financial crisis, along with LTCM episode and October 1987 stock market crash, helped achieving some degree of consensus regarding the call for models or techniques able to properly and efficiently detecting and assessing systemic risk. In this sense, taking into account that financial markets are a complex system, where connections matter as much as the participants that originate them, TCTF concept has emerged along with the traditional TBTF concept. Some authors agreeing with this statement are Chan-Lau (2010), ECB (2010), Clark (2010) and Zhou (2009).

Akin to the TBTF concept, TCTF could be straightforwardly defined. Based on a customary definition of systemic risk (CPSS, 2001), an institution may be labeled as TCTF when, due to its connectedness – either direct or indirect –, its inability to meet its obligations could result in the inability of other system participants of the financial system to meet their obligations as they become due.

Nevertheless, identifying a TCTF institution is not as straightforward as with a TBTF, where the latter relies on observable concepts and metrics such as the assets' value or the volume of financial services (eg deposits and loans) provided by a single institution. Identifying a TCTF institution is intricate. Among the main sources of intricacy it is worthwhile stressing that measuring an institution's connectedness is (i) complex and cross-dependent by nature (ie it cannot be measured in isolation) and (ii) extremely context-dependent and non-linear (Landau, 2009, Haldane, 2009).

Financial markets constitute one among many other systems exhibiting a complex organization and dynamics, where the large number of mutually interacting parts self-organize their internal structure and dynamics with novel and sometimes surprising macroscopic emergent properties (Sornette, 2003). Such surprising emerging properties include, for instance, a tradeoff between risk sharing and systemic risk, where the degree of connectivity increases above a certain threshold, crises tend to be not only more severe, but also more frequent (Battiston et al, 2009). Or, as suggested by Trichet (2009), complexity makes that what in tranquil times is an efficient mechanism to share risk, can, in times of stress, become a dangerous channel for transmitting instability.

Regarding the first source of intricacy for identifying a TCTF institution (ie complexity and cross-dependency), standard econometric approaches are not suitable for the task because the properties and behavior of the institution cannot be analyzed on the basis of its own properties and behavior alone, as these may be affected by institutions that have links to it, and also by other institutions that have no direct links, but are linked to its neighbors

(ECB, 2010). A key concept when defining whether an institution fits the TCTF concept or not is centrality. Akin to Schmitz and Puhr (2007) study of the Australian large-value payments (ARTIS), this paper embraces degree centrality as a metric for connectedness, where its dominance over betweenness centrality for capturing systemic importance is its main advantage.

About context-dependency, because of the numerous participants and connections within the financial market, a minor change in the initial conditions of the system (eg liquidity, regulation, macroeconomic environment) may critically affect the result of the analysis. This is, the TCTF label is particularly dynamic, where an institution may become TCTF (or non-TCTF) with an otherwise trivial alteration of the status-quo. Sensitivity to initial conditions is intuitive in the financial markets' case since, for example, abundant liquidity allows institutions to process payments independently from each other, whilst reduced liquidity makes institutions' ability to make payments become coupled with the ability of other institutions to make payments (Kambhu et al, 2007).

Similarly, as with context-dependence, the non-linearity features of complex systems set additional hurdles for identifying TCTF institutions. Non-linearity refers to the disproportionate effect of a shock in the overall properties of the system, which are not easily captured by standard econometric tools. For the subject under study, for example, there is some degree of consensus about the lack of correspondence between the subprime crisis (ie the shock) and the global financial crisis (ie the outcome), where the former is rather modest when compared to the extent of the whole episode (Bullard et al, 2009, Haldane, 2009, Gai and Kapadia, 2010).

Regarding these two sources of intricacy, Natural and Social Sciences have faced similar challenges. For the first one (ie complexity and cross-dependency), network topology (also referred as network analysis or theory, NT) has been a widely accepted and highly convenient approach, since it allows for analyzing systemic risk by looking at how resilient the system is to contagion and what the major triggers and channels of contagions are (Tumpel-Gugerell, 2009), taking into account the size of flows, interdependencies with other systems/markets, and the degree of substitutability (Manning et al, 2009).

For the second source of intricacy (context-dependence and non-linearity), simulation models are useful tools because they can be calibrated to replicate a specific environment (Arjani et al, 2007), and because it allows for assessing the impact of altered liquidity levels and payment flows in terms of payment queues, liquidity requirements (eg overnight lending) and the value and number of unsettled transactions (Leinonen and Soramäki, 2005).

At the end, because the system is robust-yet-fragile-and-uncertain due to its complexity, homogeneity and opaqueness, where the existence of context-dependency and non-linearity further hinders the use of typical approaches to systemic risk, the use of a combination of network theory and simulation techniques may be a comprehensive and convenient framework for identifying and assessing systemic risk.

11.3 Colombia's payments system's stability under the too-connected-to-fail concept

A payments system (PS) is a set of instruments, procedures and norms for the transfer of funds among participants in the system (CPSS, 2001). Accordingly, the effectiveness and stability of financial markets depends on it functioning properly. Considering how important settlement and large-value payment systems are to financial stability, the central banks in most countries now own and operate these systemically important systems, which use real-time gross settlement (RTGS) as the primary method of settlement.[1]

The systemic importance of a PS depends on the functions it fulfills within the economy. According to CPSS (2001), a systemically important PS has at least one of the following characteristics: (i) it is the only PS in the country or the main PS in terms of the aggregate value of the payments; (ii) it primarily handles payments of large individual value; and (iii) it is used to settle financial market transactions or to settle other payment systems.

This chapter is dedicated to introducing the approach developed by Banco de la República (Machado et al, 2010) for identifying and assessing systemic risk within Colombia's financial market. The approach is based on an application of NT and simulation techniques to BR's large-value PS (henceforth referred as CUD), which is the only large-value PS in the country and is used to settle all financial

[1] The RTGS mechanism is characteristic of payments systems managed by central banks, where clearing and settlement are processed immediately and simultaneously in the accounts the institutions have with the central bank. In 2008, the World Bank surveyed 142 central banks about their PS and found that 112 of the large-value PS settle their operations according to the RTGS scheme and 108 of these PS (96%) are operated by the central bank (World Bank, 2009).

market transactions and to settle other PS, thus a systemically important SP by CPSS's (2001) standards.

Different from other approaches based on network topology, the chosen approach relies on payments as connections. Balance sheet claims between institutions as sources of connectedness (Chan-Lau, 2010b) are deemed by the authors as impractical for the purpose of this document because (i) it is not clear whether off-balance positions are being captured or not when using claims, whilst payments comprise all transactions between payments system's participants; (ii) unlike claims, relying on payments allow for considering liquidity as a key factor in systemic risk; (iii) as acknowledged by Tumpel-Gugerell (2009), a particular institution might not only be systemically relevant because other institutions are financially exposed to it, but also because other market participants rely on the continued provision of its services; and (iv) as emphasized by Kodres (2009), failure or insolvency are not the only sources of systemic shocks, but mere failure-to-pay or non-payment of transactions can gridlock the entire financial system.

11.3.1 A brief introduction to Colombian large-value payments system

The Colombian PS comprises a centralized network infrastructure in which BR's CUD operates as a hub that maintains communication with all other participants (ie securities depositaries, low-value payment systems, the Foreign Exchange Clearing House, the Chamber of Central Counterparty Risk), where the participant that generates most activity and volume is BR's own securities depositary (DCV), which is exclusively dedicated to clearing and settlement of the most liquid fixed-income securities in the local market: which is central government's local public debt bonds called TES.[2] All financial institutions – and some special official entities – are allowed to participate directly in the CUD, which is in charge of the clearing and settlement of all their payments; the most relevant types of financial institutions participating in CUD are briefly described in Table 11.1.

The CUD started in September 1998 and, since then, it has operated as a RTGS system, with its monthly volume representing as

[2] Central government's local public debt bonds (TES) are the most liquid securities in the local market. They correspond to the most used eligible collateral for accessing central bank's liquidity.

much as 1.71 times the GDP (September and October 2009). During 2009, 160 institutions directly conducted transactions in the CUD, where Commercial Banks (CB) and Brokerage Firms (BF) were the most active with about 75% of all operations.

Table 11.1 **Main Colombian market's financial institutions directly participating in CUD (2009)**

Class	Institution type	Main purpose[c]
Credit institutions (CI)[a]	Commercial Bank (CB)	Provision of deposit and loans, including mortgages. [18]
	Commercial Financial Corporation (CFC)	Provision of deposit and loans focused on goods and services commercialization (eg leasing). [26]
	Financial Corporation (CF)	Provision of deposit and loans focused on medium term industrial financing; akin to an investment bank. [3]
Non-credit institutitons (NCI)	Mutual Fund (MF)	Provision of investment vehicles with the purpose of investing in securities and other assets according to the risk profile of the investor. [26]
	Brokerage Firm (BF)	Provision of brokerage services with the purpose of buying and selling securities (eg stocks, bonds, currencies); allowed to trade for its own account. [32]
	Pension Fund Manager (PFM)	Provision of investment vehicles with the purpose of investing for retirement. [6]
	Special Official Institution (SOI)	Official (government owned) financial institutions with special objectives; due to its main features, they were excluded from the analysis.[b] [10]

[a] Financial cooperatives pertain to Credit Institutions, but due to its low connectedness and size they were excluded from the analysis; CIs are the only institutions able to receive LLR liquidity.
[b] SOI type comprise ten government owned institutions, where the largest is Fogafin, the deposit insurance agency. Their involvement in the CUD is rather low, thus they were excluded from the analysis.
[c] Only the main differencing feature appears; the number of institutions as of 2009 appears in brackets.
Source: authors' design.

As previously mentioned, the CUD, unlike other countries' PS (eg CHAPS Sterling in the U.K.), is a direct participation system where any type of financial institution can maintain deposits and conduct transactions with other participants without the need for an agent or intermediary. For this reason, the CUD has a large number of direct participants (160) representing all types of institutions, banking and non-banking.

When institutions participating in the PS experience temporary liquidity problems, they can make use of BR's resources through different facilities. Within its expanded inflation targeting scheme, in which the stability of the financial system plays an essential role, BR's liquidity facilities can be grouped according to their objective, namely: (i) for macroeconomic liquidity, through Open Market Operations (OMO); (ii) for the ordinary operation of the PS, through intra-day repos, which may be converted into overnight repos; and (iii) for financial stability, which is achieved when BR fulfills its LLR function.

Regarding OMO, they are BR's main monetary policy instrument, as is the case with most central banks that use an inflation targeting approach. OMO transactions (via selling or purchasing TES) are conducted by OMO agents, which by the end of May 2010 accounted for 97 institutions, where CBs, CFCs and CFs are the most active.

About the second facility, BR introduced two instruments to complement OMO and to ease PS's liquidity pressures: the intra-day repo and the overnight repo. Intra-day repos first became available in 1999 and are used by institutions to cover their liquidity shortages during the trading hours. The overnight repo facility has been in place since 2001, and it materializes in two ways: (i) after an institution fails to fulfill an intra-day repo, and (ii) when a CB does not have enough funds to clear checks.

Concerning the third facility, BR can act as a LLR to minimize contagion and to keep the financial system stable. As asserted by Meltzer (1986), under special conditions, this function allows central banks for providing the resources an institution needs to deal with a transitory liquidity problem. In Colombia this is known as Transitory Liquidity Facility (TLF), and is reserved exclusively for Credit Institutions, which are firms dedicated to the provision of deposit and loan products, namely CBs, CFCs and CFs.

11.3.2 Network topology and payments simulation for identifying and assessing systemic risk

The periods and the institutions to be evaluated and analyzed were defined in order to assess the systemic risk and potential threat to the stability of the PS and the financial markets. Three periods representative of CUD transactions were selected based on the concept of liquidity and TES market activity, which follows the need for effectively capturing different volatility and liquidity scenarios for the Colombian financial market, allowing for a better assessment of the dynamics of PS stability. Afterwards, based on the selected scenarios, four institutions were selected within the CUD based on a first approach to systemic importance based on degree centrality, as will be explained later in this document.[3]

Pursuant to the foregoing, the month of June 2006 was selected as the period representative of high volatility in the PS, corresponding to the most recent acute stress in the TES market; akin to 1987's crash, LTCM's and the most recent global financial crisis, June 2006 was characterized by a run on local market liquidity, where the non-banking financial institutions, namely BFs and MFs, were particularly threatened. This period witnessed a sharp drop in the price of TES (local market's benchmark).[4]

Contrasting with June 2006, four months before was a period characterized by high liquidity and low volatility in Colombian financial markets, where the TES market exhibited the peak of a prominent boom. Thus, February 2006 was chosen as representative of tranquil times, with abundant liquidity and confidence among market participants.

September 2009, the month when the CUD registered the largest trading volume since its creation, was selected as well. In all, 215,776 transactions were conducted during that month for a volume representing 170.7% of GDP, with a daily average of 9,808 transactions.

The NT and payment simulation is applied to a single day of transactions. There are two main alternatives for defining the single day to be used by the model. First, an observed single day of transactions (eg February 1st, 2006) can be used, with the advantage of preserving the true dynamic and serial dependence of transactions, but with potential biases resulting from particular transactions that are infrequent but significant for affecting intraday liquidity in a concealed manner (eg central government's securities interest and principal payments, social security transfers, taxes collection).

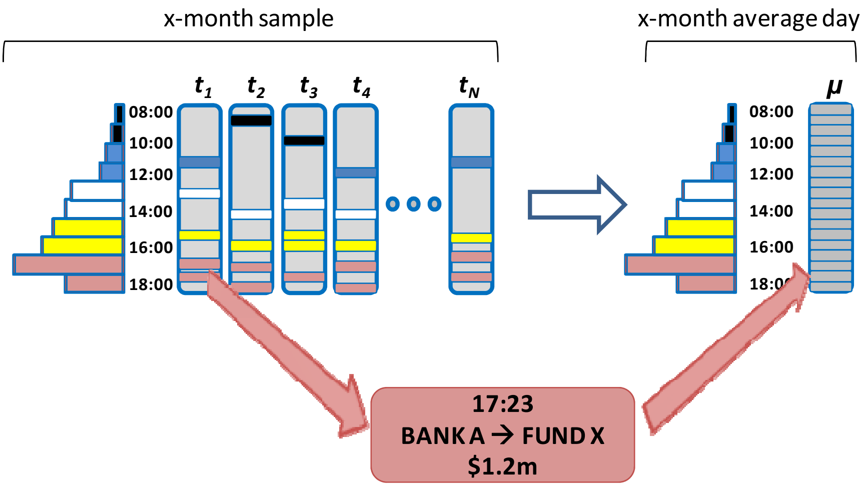
---

[3] According to authors' calculations, highly connected (central) financial institutions correspond to those that affect the most the system as a whole. Selecting the four institutions of greatest systemic importance follows practical reasons.

[4] Based on the fixed-income index (IDXTES) developed by Reveiz and León (2008), June 2006 TES returns reached –2.26% (–8.4 standard deviations).

Second, a synthetic typical day of operations can be used. The construction of such a typical day of transactions may be done by using the bootstrap method. A month of intraday transactions' database is divided in several time-of-the-day buckets (eg 7:00–7:59, 8:00–8:59 … 20:00). A traditional resampling with replacement procedure (Dowd, 2002) is employed to sample transactions from each bucket until reaching the average volume of transactions for each bucket and the overall volume of transactions per day for the selected month (Figure 11.1). Under this approach the resulting synthetic typical day of transactions fairly approximates the main distributional moments (eg mean, variance, skewness and kurtosis) of the overall transactions for the month of intraday transaction's database, whilst capturing the average intraday dependence between the different time-of-the-day buckets.[5]

Figure 11.1 **Estimating an 'average day' of payments transactions**



Source: authors' design.

As expected, the main drawback of this approach is to average out the serial dependence of transactions, where intraday dependence is

[5] Standard statistical tests (eg Kolmogorov–Smirnov) confirmed that the distribution of the synthetic typical day of transactions does not deviate significantly from the overall transactions for the three months of intraday transaction's database herein modeled (February and June 2006, and September 2009).

smoothened but not ignored. The advantages of this approach are (i) to avoid the biases aforementioned and (ii) to be able to characterize intraday payments dynamics for a whole period that is well-known and valuable for its informational content (eg high volatility, market boom, etc.).

Because being able to use periods distinguished by their informational content is key for the purpose of this paper, along with the convenience of avoiding the aforementioned bias resulting from a single observed day of transactions, the second alternative (synthetic typical day of transactions by bootstrap) is employed. Using an average day estimated in this way, rather than any particular single day, allows for a more robust characterization of the stability of the PS network in different scenarios, as it preserves the conditions found during the period (ie liquidity, intra-day seasonality), while mitigating the impact of infrequent but large operations on the part of certain non-financial participants (eg Ministry of Finance). However, this step may be skipped and a single day of data may be used if deemed appropriate.

As a primary approximation to the notion of systemic risk on the average day of transactions for each period an index based on the idea of TCTF was constructed. The overall index takes into account (i) each institution's share of the total traded value and (ii) each institution's share of the total number of connections during the three scenarios;[6] these two measures are traditional (in-degree and out-degree) measures of centrality, a concept which refers to the importance and location of the participant or node in the network (ECB, 2010, Soramäki et al, 2007), and – as previously mentioned – is the most appropriate measure to this kind of network of liquidity flows (Schmitz and Puhr, 2007).

This index serves as a primary approximation to the notion of systemic risk based on the idea of TCTF. Table 11.2 shows the results of the index for the ten foremost connected (central) institutions. These ten institutions, which represent 6.3% of the CUD participants, account for 47.4% of the traded value and 25.8% of the connections. Among the top ten institutions of major systemic importance

[6] In this index, the institution with the largest share of total value or total connections obtains a score of 100. The following institutions, by linear interpolation, obtain a score between 0 and 100. The aggregate corresponds to the sum of the index obtained for each institution in both categories, which is then used to calculate a general or overall index. Participants such as the Ministry of Finance were excluded from this analysis; their characteristics demand a special study to assess their systemic impact.

according to the centrality concept there are seven CBs, two BFs and one CF.

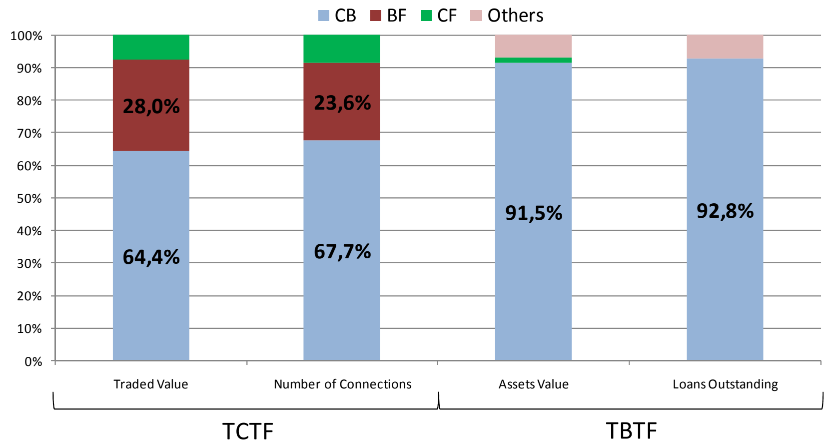
Table 11.2 **Ten foremost connected (central) institutions[a]**
3-period average

	Traded value		**Number of connections**		**Aggregate**	
Institution	**Share**	**Index [A]**	**Share**	**Index [B]**	**[A+B]**	**Overall index**
INST1	8,4%	100	3,5%	100	200	100
INST2	7,1%	85	3,1%	87	139	70
INST3	6,6%	78	2,8%	78	127	63
INST4	4,9%	58	2,6%	72	107	53
INST5	5,0%	60	2,5%	70	106	53
INST6	3,7%	44	2,7	77	101	50
INST7	2,3%	27	2,7%	77	89	44
INST8	3,6%	43	2,2%	63	88	44
INST9	2,8%	33	2,0%	55	74	37
INST10	3,1%	36	1,8%	51	72	36

[a] The ten foremost connected (central) institutions represent 47.4% of the total traded value and 25.8% of the total connections.
Source: authors´ calculations.

As indicated in Figure 11.2, pursuant to the notion of TCTF (left side of Figure 11.2), the CBs are the institutions of greatest systemic importance within the ten foremost connected institutions (64.4% and 67.7% of traded value and number of connections, respectively), although the BFs are significant as well (28.0% and 23.6%, respectively).

Figure 11.2 **Ten foremost relevant institutions: TCTF and TBTF 3-period average**



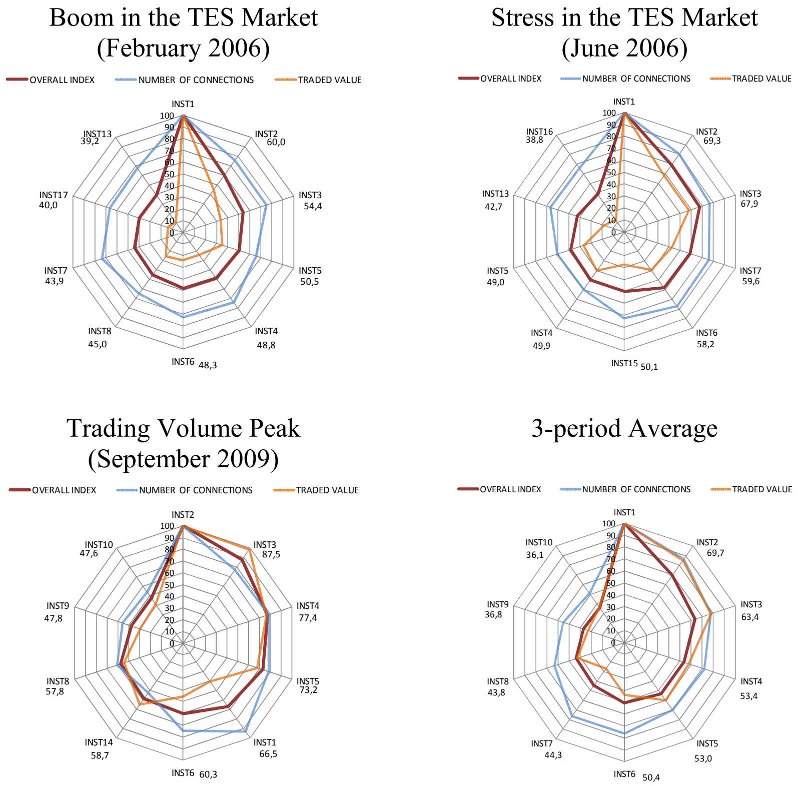
Source: authors' calculations.

Pursuant to the TBTF concept, which is customarily measured according to the asset value and the loans outstanding[7] (right side of Figure 11.2), CBs are practically the only institutions that may be regarded as systemically relevant (91.5% and 92.8% of assets value and loans outstanding); the remaining share of asset value loans outstanding pertains to Fogafin (the deposit insurance agency) and other SOI. This stresses the importance of considering connectivity as a measure of systemic risk.

Figure 11.3 portrays the make-up of the index of systemic importance for the top ten institutions in the three periods and an average of these. The average's first seven institutions appear in all the selected periods. This suggests the institutions with more systemic risk, pursuant to the notion of centrality, are relatively stable over time.

[7] Another customary measure of size is the value of deposits. It is not included because CB, CF and CFC are the only authorized institutions to take deposits from the public.

Figure 11.3 **Ten foremost relevant institutions: TCTF and TBTF**



Source: authors' calculations.

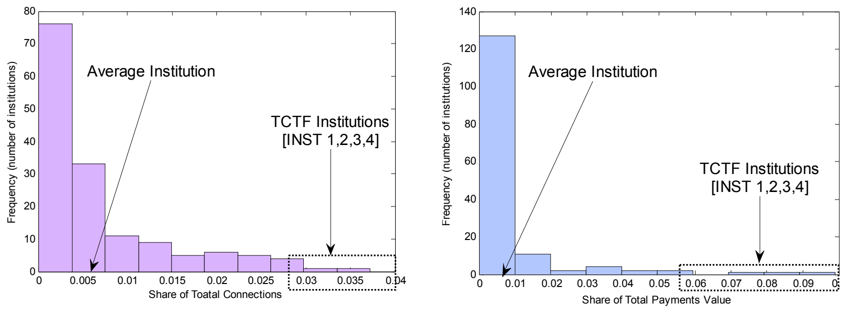
The first four institutions of greatest systemic average importance according to the centrality metrics were selected to carry out targeted shocks (attacks) on the PS (ie INST1, INST2, INST3, and INST4 in Table 11.2).[8] The result was a sample with two types of institutions: CBs and BFs. The systemic effect of an attack on each of the four selected institutions will be simulated in the following sections, and

[8] As previously mentioned, according to authors' calculations, highly connected (central) financial institutions correspond to those that affect the most the system as a whole. Selecting the four institutions of greatest systemic importance follows practical reasons.

the results will be presented and analyzed as an average per type of institution.[9]

Figure 11.4 **Distribution of connections and payments per institution (%)**
3-period average

Connections (in and out) Payments (value)



Source: authors' calculation

It is worthwhile to emphasize the importance of the usage of centrality for deciding which institutions will be attacked in the NT and simulation approaches. As documented before, because the distribution of institutions' connectivity is significantly fat-tailed and skewed (Figure 11.4), using random shocks will tend to overlook systemic risk; TCTF entities await in the uttermost right side of the distribution. Assessing the centrality of institutions allows for selecting those entities that actually may endanger systemic stability, whilst preserving parsimony within the approach.[10]

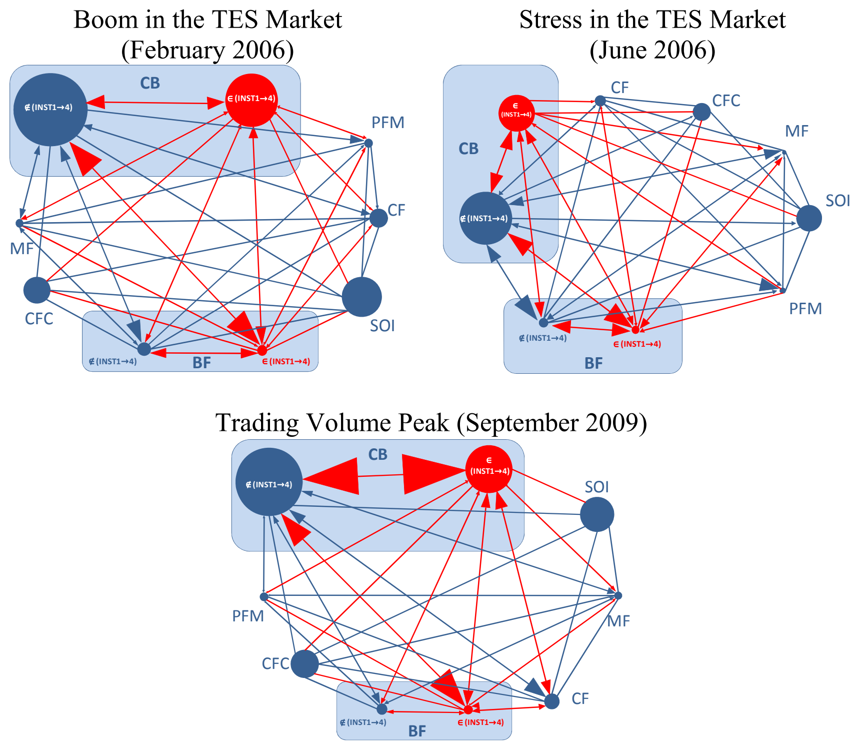
As before (Figure 11.2), using centrality as an objective metric for systemic importance based on the TCTF concept allowed for identifying institutions which would have been disregarded by the traditional TBTF concept. This is evident in Figure 5, where network theory (graphs) is used to simultaneously present TCTF and the TBTF

[9] Authors deem necessary to present results as an average per type of institution (eg CB and BF) in order to preserve confidentiality. This is also the reason not to disclose the names of the institutions or the precise composition of the top-four and top-ten systemic relevant firms by type of institutions.

[10] Instead of using random shocks or targeted shocks based on centrality it is possible to shock all the institutions of the system. Nevertheless, this may be computationally inefficient and burdensome, especially in a PS where any financial institution may participate directly; this is the case in hand, where it would be necessary to shock, simulate and analyze 160 institutions.

concepts for each scenario. The size of each arrow's head represents the total value of the payments (ie connectedness), whilst the size of the node represents the asset size; please note that the four foremost relevant institutions under the centrality concept and their connections have been differentiated (in red) for illustrative purposes, where the [∈(INST1→4)] nodes corresponds to the institutions belonging to INST1, INST2, INST3 or INST4, and the [∉(INST1→4)] to those not belonging, either pertaining to the CB or BF institution types (in rectangles).

Figure 11.5 **Colombia's financial system as a network (graphs)**



Source: authors´ calculations.

Regarding the graphs in Figure 11.5 it is key to emphasize the following: (i) as expected, focusing on the size of institutions (nodes' size) would concentrate supervision on Credit Institutions (CBs, CFCs and CFs), along with government related SOIs; (ii) focusing on the size of the institutions would overlook the importance of the 'shadow banking system', with BFs being institutions heavily connected to CB for the three scenarios considered; (iii) the nodes corresponding to the four foremost systemic relevant institutions [$\in$(INST1$\rightarrow$4)] concentrate a significant share of the connections within the network, thus supporting their choice for illustrative purposes.

All this further supports the mainstay of the TCTF concept: using the size of the institution is insufficient to assess its contribution to systemic risk within the financial network; it is necessary to shift from micro-prudential to macro-prudential approaches, where the latter refers to the objective of promoting the stability of the financial system as a whole (Clement, 2010).[11]

11.3.2.1 The static approach: network topology

Network topology (also known as network analysis or network theory) is a method used in statistical physics to understand and analyze the structure and functioning of complex networks. Recent applications demonstrate its usefulness in analyzing how PSs respond to liquidity stress. The studies by Soramäki et al (2006) and Bech and Garrat (2006) use NT to characterize the PS in the United States (Fedwire),[12] while Ianoka et al (2004) apply it to the PS in Japan (BoJ-Net). In the Colombian case, Cepeda (2008) applies NT to the CUD to quantify the impact of failures on its stability.

A PS network is made up of a set of nodes or vertexes (institutions participating in CUD) and connections or links (payments) between pairs of nodes. The connections between nodes can be directed or not, and can be weighted (by value or volume of payments) to reflect the strength or weight of the link that is established. Accordingly, the PS is constructed of nodes or participants, which are the institutions that comprise the network and conduct transactions with one another. Based on this characterization, it is possible to study the basic properties of the network, which are observed through parameters such as average distance, diameter and connectivity.[13]

Calculating the stability of the PS based on NT is intended to characterize the CUD as a network, so as to estimate its stability in

[11] Central bank's macro-prudential role makes part of its oversight function, which has the objective of attaining an efficient and safe payments and settlement system (CPSS, 2005).

[12] The applications of Lublóy (2006) for Hungary and Boss et al (2004) to measure systemic risk in the Austrian banking system are useful examples of the use of this method to characterize the banking system.

[13] Basic concepts of NT applied to PS's stability analysis are used in this paper. Cepeda (2008) describes and uses additional concepts and metrics.

scenarios that involve an institution's failure-to-pay or non-payment. Initially, the parameters were calculated for the CUD payment network in each of the selected periods (average days for February and June 2006, and September 2009). The next step was to recalculate the network's parameters after subjecting it to failure-to-pay by a selected node; this is a targeted shock or attack. This was done by eliminating the transactions originating from that institution (out-degree), but preserving those made by its counterparties (in-degree), including transactions which correspond to payments directed to the failing node.

Four failure-to-pay or non-payment scenarios were designed for each period, with each corresponding to an institution that ceases – for whatever reason[14] – to make its payments (ie INST1, INST2, INST3, INST4 from Table 11.2). These attack scenarios are interpreted as static in nature, since they consider all the transactions sent/received in the selected average day regardless of the order in which they were conducted, nor taking into account whether the participants had enough funds on their BR's deposit accounts to fulfill those payments. Assuming that institutions had enough funds may result in some nodes or participants remaining 'artificially' connected to the network after the attack because they are taken as able to comply with their payments when, in certain cases, they could have exhausted their liquidity.

Afterwards, the approach consisted of evaluating the change in network's parameters (ie connectivity, diameter, average distance), identifying the type of network (eg the distribution of number of connections per node), as well as measuring the effect caused by the attacked institution on the traded amount and the number of disconnected institutions. As asserted by Becher et al (2008), if the network is robust and stable, the attack should have little effect on other participants; that is, the properties of the network should not differ significantly compared to those in the original scenario (with no attack).

The results of the exercise are exhibited in Table 11.3, in the form of variations with respect to the original scenario. It is evident that the impact on the network varies depending on the type of institution under attack and the selected period. For example, the attack on BFs has more of an impact on the network during the boom and stress

[14] It is important to emphasize that the reason behind the non-payment or failure-to-pay of the institution is non-specified; it may be due to liquidity problems, solvency, operational risk, legal risk, etc.

periods, while the CBs had the most impact during the period of greatest activity in the PS.

Table 11.3 **Attacks' effects on the network (CUD) – static approach (NT)**
As variations with respect to the original scenario – per type of institution and scenario

Scenario	Institution type	Network activity		Network topology criteria						
				Network features			Nodes features[a]			
		Traded value	Number of transactions	Distance	Diameter	Connectivity	Sending & receiving	Sending only	Receiving only	Not receiving & not sending
TES boom (Feb/2006)	CB	–5,9%	–4,7%	0,01	0	–2,8%	–1	1	0	0
	BF	–8,5%	–7,4%	0,01	0	–4,5%	–1	0	1	0
TES stress (Jun/2006)	CB	–6,5%	–4,9%	0,03	1	–2,3%	–1	0	1	0
	BF	–7,1%	–7,2%	0,01	0	–3,3%	–1	1	0	1
CUD trading peak (Sep/09)	CB	–9,1%	–6,6%	0,01	0	–3,3%	–2	1	1	0
	BF	–5,4%	–5,3%	0,02	0	–2,6%	–1	1	0	0

[a] Corresponds to the number of nodes; rounded to the next integer.
Source: authors' calculations.

The attack on the BFs in the boom scenario (February 2006) had a larger average impact in terms of reducing the amount traded (8.5%) than the impact derived from the attack on the CBs (5.9%), which was also reflected in the decline in the number of transactions (7.4% and 4.7%, respectively). During that same scenario, attacks on the institutions caused no major changes in distance and diameter, although connectivity declined considerably.

Failure-to-pay by BFs in the stress scenario (June 2006) generated less of a reduction in the amount traded and the number of transactions than was the case during the previous scenario (7.1% and 7.2%, respectively), and a lower connectivity loss.[15] It is noteworthy that failure-to-pay by BFs generates a marginal loss in network stability, but does cause one participating institution to disconnect

[15] As presented in Machado et al. (2010) the stress period (June 2006) exhibits the lowest level of connectivity among the chosen scenarios (60.6%, 46.1% and 49.7% for scenarios 1, 2 and 3, respectively), which intuitively results from participants' reluctance to engage in market transactions. This explains why the connectivity loss during the stress period is – on average – the lowest.

from the network; in other words, it neither receives nor makes payments within the PS.

During the period of increased activity in the network (September 2009) the average failure-to-pay by CBs results in a decline of 9.1% in traded value within the PS and 6.6% in terms of the number of transactions. The distance increases by 0.01, the diameter remains the same, and connectivity is down by 3.3%. Contrary to the previous period, there were no disconnected nodes.

The average result of the attacks during the three selected periods is shown in Table 11.4, by type of institution. CBs are the institutions with more of a direct impact on the volume traded, since their failure-to-pay would lower the average traded amount in the CUD by 7.2%, as opposed to a decrease of 7.0% by BFs.

Table 11.4 **Attacks' effects on the network (CUD) – static approach (NT)**
As variations with respect to the original scenario – per type of institution

Institution type	Network activity		Network topology criteria						
			Network features			Nodes (institution) features[a]			
	Traded value	Number of transactions	Distance	Diameter	Connectivity	Sending & receiving	Sending only	Receiving only	Not receiving & not sending
CB	–7,2%	–5,4%	0,02	0	–2,8%	–1	1	0	0
BF	–7,0%	–6,6%	0,01	0	–3,5%	–1	0	0	0

[a] Corresponds to the number of nodes; rounded to the next integer.
Source: authors' calculations.

Nevertheless, the SBFs have more of a direct impact in terms of the average number of transactions, which are down 6.6% compared to the reduction of 5.4% generated by the attack on the CBs. As for network connectivity, the attack on the BFs results in an average decline of 3.5%, which is more than the reduction caused by the attack on the CBs (2.8%).

Network topology confirms that the TBTF concept is insufficient to identify systemic risk sources. For Colombian large-value payments system (ie CUD), focusing on the size of the institutions would result in overlooking the importance of BF, which appear to be of similar systemic relevance as CBs.

11.3.2.2 The dynamic approach: payments simulation

NT can be used to characterize and analyze the structure and operation of complex networks. However, in its basic form, this approach can have certain limitations. As presented so far, the attack involved removing a node or participant as the originator of transactions under two key assumptions: (i) other institutions are always able to fulfill their obligations (ie their intra-day liquidity level is not considered) and (ii) other institutions do not react to the attack.

Relaxing the first assumption involves acknowledging that the capacity of institutions to conduct their transactions depends on their opening balance in the CUD at the start of the day, as well as all the transactions that imply an inflow and outflow of resources for them during the day, where the opportunity of each transaction is related to the ability of each institution to fulfill its obligations. On the other hand, relaxing the second assumption would demand making additional and challenging assumptions about how information spreads throughout the financial markets and about the manner non-attacked institutions react upon the arrival of this information; this is the reason why this assumption was preserved.[16]

Simulation exercises may provide the central bank with additional information that is valuable for managing liquidity in the PS. In this respect, Leinonen and Soramäki (2004) suggest that simulation analysis of PS transactions makes it possible, among other things, (i) to quantify the result of a change in payment flows; (ii) to determine the result in payment queues and liquidity requirements owing to the change in payment flows; and (iii) to quantify the need for overnight liquidity or the value and number of transactions that would not be completed if additional liquidity is not available.

---

[16] Relaxing this second assumption requires a more extensive study to identify information conditions in the market (eg the existence of asymmetries, intra-group information management, and the response strategy by type of institution or particular entity), which is beyond the scope of this work; this assumption is common when using simulation techniques (Leinonen and Soramäki, 2004), but has been addressed by Soramäki et al. (2007). Authors consider this assumption as a rather interesting starting point for two reasons. First, it may be regarded as a stringent case in which a system's hub will act as a liquidity drain, where a major participant receives all their counterparties' payments but makes no payments to its counterparties. Second, because each source of systemic shock entails different informational dissemination and reaction dynamics (eg an operational driven failure may become noticeable later than a solvency issue), maintaining the non-reaction assumption serves the purpose of not specifying the source of shock. Nevertheless, authors acknowledge that the failure of a hub becoming public could have major consequences for the institutions' willingness to make payments to each other or even bank runs, which may generate an extreme disruption case.

In order to get information not available through traditional statistical or econometric approaches neither to NT alone, the methodological approach used in this analysis seeks capitalize on benefits of NT and simulation in payments systems to develop a dynamic analysis that measures the direct and indirect impact of the attack.

This approach, which compares end-of-the-day liquidity in a base scenario (ie with no attack) to end-of-the-day liquidity in a scenario where there was an attack on an institution, captures the direct and indirect effects of the failure-to-pay, the latter being caused by having connections with a previously affected institution; this is, this approach captures the attack's extended (ie second-hand) effects, which are the mainstay of systemic risk. Afterwards, based on NT, the simulation of payments' results are used to compare the network's properties (eg connectivity, diameter, average distance) before and after the attack. Finally, the responsiveness and resilience of participating institutions was analyzed according to their financial structure and access to BR's liquidity in the event of attacks on the PS.

The simulation of payments uses the opening balances and CUD transactions for each of the three selected periods as the base scenario, where central bank's liquidity facilities is excluded from CUD's transactions.[17] The base scenarios are compared to the attack scenarios, in which failure to fulfill any payment by one of the four institutions selected according to their degree of connectivity is assumed to occur from 9:00a.m onwards. The parameter for comparison will be the variation between the base scenario and the attack scenario with respect to the unresolved payments each institution still has on queue (Payments on Queue, PoQ) at the end of the day.

The simulation is based on the liquidity institutions have in their deposit accounts (opening balance), which is affected during the day by the transactions registered chronologically in the CUD. Because the objective is to replicate a RTGS large-value PS (ie CUD), a transaction can be carried out only if the institution making the payment has enough funds in its deposit account. If it does not, the payment is placed on the queue of outstanding payments (PoQ). PoQ will be fulfilled to the extent the institution obtains enough funds to

[17] Excluding central bank's liquidity facilities from the central bank is convenient because it allows for obtaining true end-of-the-day liquidity requirements by the participating institutions, thus allowing for assessing the total liquidity required by the system and each participating institution. This will also allow for properly evaluating the sufficiency of central bank's liquidity facilities.

cover all or part of them, for which a Queue Resolution Algorithm (QRA) was defined.

The selected QRA is based on the 'First In First Out' or FIFO algorithm, which is the one most commonly used.[18] Every time a transaction is executed (ie a payment is made), the algorithm assesses if the institution that received funds has any PoQ and whether or not the new balance is sufficient to cover any of these left-pending transactions. The algorithm respects the order in which the unfulfilled transactions were placed on queue. If it is possible to settle one of the outstanding payments, it is registered as a new transaction. This, in turn, prompts the QRA to search again to determine if the institution that received funds has any PoQ, if the new balance is sufficient to settle any of the outstanding payments, and so forth.[19] This type of QRA is known as FIFO-bypass (BR, 2009).

To manage the liquidity in the system more efficiently a multilateral clearing and settlement algorithm was used to net institutions' PoQ at different points in time. For this purpose five multilateral nettings were carried out during the trading day, all made during periods of peak intra-day trading. Liquidity savings are generated this way, giving participating institutions more capacity for settling PoQ and allowing for more efficient use of liquidity in the simulation.[20]

Based on the payments registered and settled using the simulation approach, the NT is applied to the simulated CUD's settled transactions with and without attack, and a new characterization of the network is provided. Table 11.5 shows a comparison of the results obtained with both approaches, where using NT on the simulations' results allows for assessing the direct and indirect (extended) impact of the attack on PS liquidity, where the intraday liquidity level is properly captured. As expected, the decline in network activity is more pronounced after using the simulation procedure, both in terms

---

[18] Leinonen and Soramäki (2004) document the existence of other types of algorithms, including those that give priority to smaller transactions and others that allow the originator of the transaction to assign a preference to each transaction; the latter is the case of the CUD in Colombia and CHAPS in the United Kingdom. The FIFO system was used because of its simplicity and given the difficulty of determining the priority assigned by each participant when registering transactions in the CUD.

[19] Jurgilas and Martin (2010) describe recent developments for managing liquidity in RTGS payment systems and implementing different algorithms to make the best possible use of the liquidity in such systems.

[20] Multilateral clearing and settlement is done pursuant to the approach used in the DCV with transactions from SEN. In the simulation procedure it is done at 12:00, 14:00, 15:00, 16:00 and 18:00 hours.

of volume traded and the number of transactions. This is an intuitive result since the simulation approach, unlike the basic NT approach, considers the direct and indirect effects of the attack and the intraday liquidity.

Table 11.5 **Attacks' effects on the network (CUD) – static approach (NT) and dynamic approach (simulation + NT)**
As variations with respect to the original scenario – per type of institution

Approach	Institution type	Network activity		Network topology criteria						
				Network features			Nodes (institution) features[a]			
		Traded value	Number of transactions	Distance	Diameter	Connectivity	Sending & receiving	Sending only	Receiving only	Not receiving & not sending
Static (NT)	CB	–7,2%	–5,4%	0,017	0	–2,8%	–1	1	0	0
	BF	–7,0%	–6,6%	0,015	0	–3,5%	–1	0	0	0
Dynamic (simul. + NT)	CB	–11,0%	–11,7%	0,019	0	–3,6%	–1	1	0	0
	BF	–12,8%	–16,8%	0,024	0	–5,0%	–1	0	0	0

[a] Corresponds to the number of nodes; rounded to the next integer.
Source: authors' calculations.

With the standard NT model CB's failure-to-pay yielded, on average, a decline of 7.2% and 5.4% in the value and number of transactions in the network, in that order. With NT applied to simulations' results, these attacks led to declines of 11% and 11.7% in value and number of transactions, respectively. Likewise, in the case of BFs results yielded a greater reduction in value (from 7% to 12.8%) and number of transactions (from 6.6% to 16.8%). As for the characteristics of the nodes, the attack on selected institutions with the dynamic model led to a sizeable reduction in activity, as well as longer distances and less connectivity; this is more evident in the BFs case.

11.3.2.3 Results: assessing systemic risk and central bank's challenges

PS activity in the event of failure-to-pay by one or more institutions can be captured through the variation in the end-of-the-day PoQ of each institution that conducted transactions during the day. By means of this analysis it is possible to: (i) identify the institutions that significantly affect the stability of the network when attacked; (ii) identify the institutions that are affected directly and indirectly as a result of the attack; (iii) quantify the impact of an attack on individual and systemic liquidity; (iv) analyze institutions' resilience, which corresponds to their capacity to deal with systemic risk by making use of their own liquid portfolios (TES), as well as BR's liquidity facilities. The results are distinguished by class and type of institution, namely (i) CIs (CB, CFC, CF) and (ii) NCIs (BF, MF, PFM), which correspond to the banking and non-banking institutions, also referred as banking and shadow-banking sectors, respectively.

Accordingly, the variation of each institution's PoQ by the end of the day allows for quantifying the liquidity shortage an institution would face, and if it has the means to absorb such shortage by using its own account TES portfolio and/or BR's liquidity facilities. With this approach, an increase in an institution's PoQ means it has been affected, inasmuch as the opening balance and the payments received from third parties were insufficient to fully meet its payment obligations, due to the failure-to-pay by the selected institution or failure-to-pay by other institutions that were affected by the failure-to-pay of the former.

No change in an institution's PoQ after the attack would mean (i) it did not cease to receive payments as a direct or indirect result of the attack or (ii) in spite of not receiving all the payments as a direct or indirect result of the attack, it had an opening balance in the CUD or a payment structure that allowed it to retain the same level of PoQ. In both cases, for the purpose of analysis, the institution is considered as non-affected.

To assess institutions' resilience, which corresponds to the magnitude of the impact on their liquidity, three ratios were designed to show the increase in PoQ as a percentage of the following variables: (i) the market value of the TES portfolio, which is the most liquid and easy-to-collateralize security in the market; (ii) the liquidity limit in OMO for each institution, and (iii) the liquidity limit in TLF for each CI.

In the case of the first indicator, the variation in PoQ with respect to the market value of the TES portfolio comes close to the concept of potential liquidity, as it would indicate whether the sale of the TES portfolio or its use as collateral would be enough to cover the outstanding payments in the CUD. A discount factor on the market value of the portfolio of each institution was used in order to capture the effect of instability on the market and its corresponding effect on securities' prices.[21] Such discount factor pertains to the maximum haircut rate used by BR (currently 3%)[22] and results in the TES Portfolio* variable.

The second ratio consists of the maximum liquidity an institution can obtain through transitory expansion operations with BR, including OMO and intra-day and overnight repos.[23] The third indicator refers to the resources CIs may access through TLF, which are the resources provided by BR on behalf of its LLR function.[24] The last two indicators make it possible to assess whether or not the current limits set for OMOs and TLF are sufficient for BR to meet liquidity needs in adverse systemic scenarios. All three ratios allow for assessing the resilience of the institutions to the attacks.

11.3.2.3.1 Scenario 1: Boom in the TES Market (February 2006)

The results, on average, demonstrates that failure-to-pay by one of the four institutions (INST1, INST2, INST3, INST4) in this scenario affects 41.3% of the CUD participants; this is, 41.3% of the participating institutions observe an increase in their PoQ. The results for the boom scenario with respect to each of the four selected institutions are presented in Figure 11.6, which relates the variation in PoQ to the market value of the TES portfolio (horizontal axis), and the variation in PoQ to the limit for accessing BR's OMO transitory

[21] It is worthwhile stating that during periods of financial turmoil (eg 2002 and 2006) the flight to quality in the Colombian market consists of shifting from TES or stocks to dollars; therefore, despite being issued by the central government and being considered as local credit risk-free instruments, TES' prices tend to fall during local crisis.

[22] This is, each institution may use (1–α) of the market value of its TES portfolio, where α is the maximum percentage of the haircut used by BR. This provides an approximate scenario to what a market stress episode may be and, in turn, yields more conservative results. The use of the 3% figure is fairly adequate since the worst daily fall ever in the IDXTES index (Reveiz and León, 2008) corresponds to 2.78% (August 22nd 2002).

[23] This limit is 35% of the liabilities subject to reserve requirements in the case of CIs, whereas for the MFs it is the value of their capital plus the legal reserve. In the case of PFMs and BFs, it is their technical capital. This document assumes that the limit allows all institutions access to the maximum amount of liquidity permitted. However, in reality, there are limits to the concentration of the auction per institution, among other constraints.

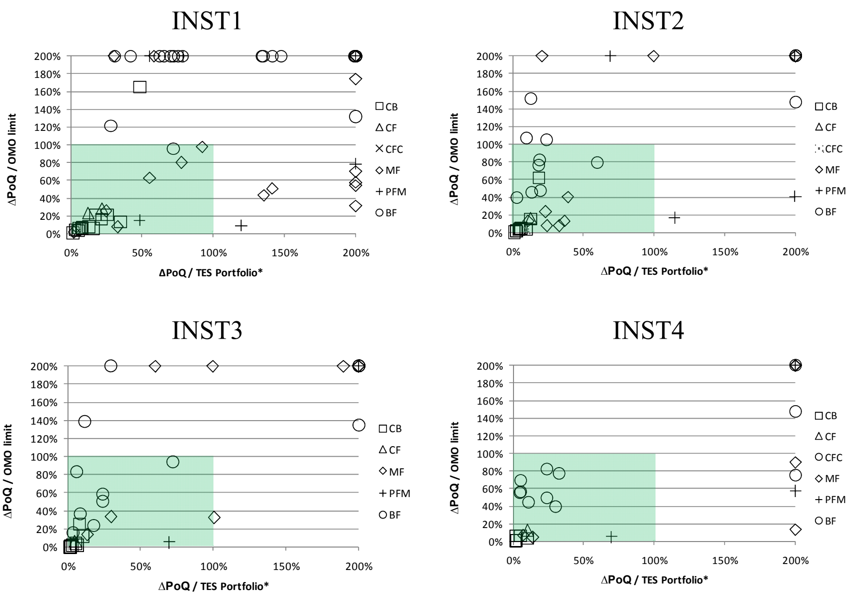
[24] The limit for TLF is 15% of the highest registered level of liabilities to the public within the 15 calendar days prior to the date the support was granted. As stipulated in the 1991 Constitution (Article 272), TLF may be used only by CIs (CB, CF and CFC); this constrain obeys the fact that CIs, due to their intermediation function (eg taking deposits and granting loans), are particularly exposed to liquidity strains.

liquidity (vertical axis); please note that both axis are truncated to 200% for practical purposes.

In this sense, the situation of the institutions in the upper right quadrant is more critical with respect to the two variables under analysis, as they would be unable to meet their liquidity needs with their TES portfolio or with OMO resources; those institutions will be referred as impacted by the failure-to-pay of the attacked institution. Meanwhile, institutions located in the shaded portion would be resilient; that is, they are able to cover the increase in their PoQ after the attacks – either by selling or collateralizing their TES portfolio or by using its OMO quota –, thus they will be regarded as non-impacted.

MFs and BFs were the institutions most affected in this scenario. INST1 was the institution that, on average, generated the most instability in the PS, having led to an increase of the PoQ for 48.5% of the participating institutions. Attacks on INST2, INST3 and INST4 affected 38%, 32% and 32% of the institutions, respectively.

Figure 11.6 **Attacks' impact on the institutions (Scenario 1)**



Note: Results truncated to 200% for practical purposes.
Source: authors' calculations.

Notwithstanding the majority of CBs and CFs witnessed an increase in their PoQ, a look at the impact of the attacks according to the type of institution shows that they had sufficient liquidity because of their substantial TES portfolio, as well as their broad access to BR's OMO mechanism; thus, CBs and CFs are examples of resilient types of institutions. Both PRMs and MFs exhibited weaknesses with respect to the use of their TES portfolio, while BFs experienced restrictions in terms of the limit on access to OMO. PFMs and MFs reduced capacity to tackle liquidity slumps is partly due to considering their TES own-account portfolio only, which corresponds to the existing regulation about restricting the use of third parties' portfolio for accessing liquidity from BR.[25]

11.3.2.3.2 Scenario 2: Stress in the TES Market (June 2006)

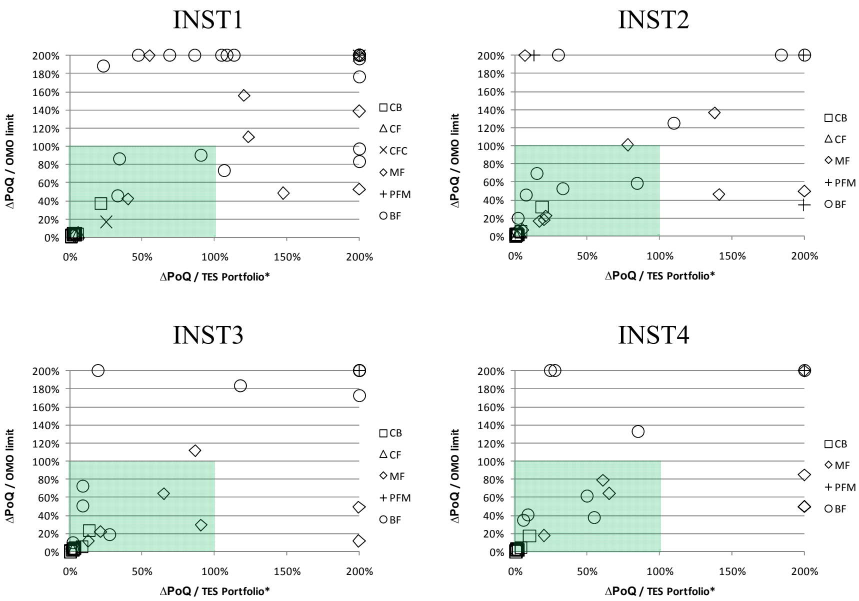
This scenario is characterized by considerable risk aversion among market participants. Unlike Scenario 1, where the TES market was booming, this scenario features a sharp decline in the mark-to-market value of local fixed income securities, which began in March 2006.

For that reason, the February 2006 and June 2006 scenarios reflect different environments in the local capital market, where TES market dropped by almost 10% in four months. This shift from a boom to a bust environment was accompanied by reduced activity in the PS. The number of transactions within the CUD fell from 9,400 to 7,377 transactions (–21.5%), whereas the traded value decreased 36.5%. Also, with respect to the characteristics of the PS in both these scenarios, the distance between nodes increased, while connectivity declined. This suggests that the system became less robust.[26]

[25] The portfolios of third parties may not be used as collateral to settle an institution's payments. This is due to regulations on separate accounting, conflict of interest and intention of the transaction between the institution and its funds. The only possibility of using them as collateral for an obligation is limited to 30% of the assets in the mutual fund and only to resolve liquidity problems specific to the portfolio in question, such as requests for withdrawals or liquidity to meet expenses.

[26] The authors found that this kind of changes in the network's properties also occurred during the transition from the first half 2002's boom to the second half 2002's uproar.

Figure 11.7 **Attacks' impact on the institutions (Scenario 2)**



Note: Results truncated to 200% for practical purposes.
Source: authors' calculations.

According to the results of the simulations, there were fewer institutions affected in the stress scenario (Figure 11.7), demonstrating the added weight exerted by factors other than the basic properties of the network. The attack on the selected institutions affected – on average – 33.9% of the participating institutions, compared to 41.3% in the boom scenario.

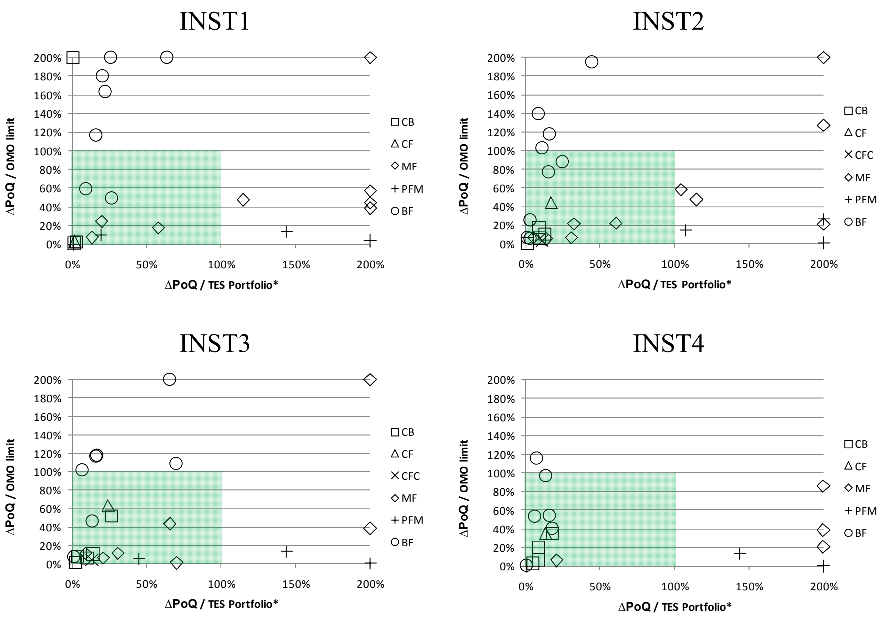
INST1 and INST2 were the institutions that affected the system the most, causing 47.5% and 31.3% of the CUD participating institutions to experience an increase in PoQ, respectively, whereas INST3 and INST4 affected nearly 30.3% and 25.2% of the participants, in that order. As for the general level of impact, despite it may seem odd to find that the consequences of an attack during stressful times are less important, participants' higher risk aversion and reluctance to engage in counterparty and market risk resulted in CUD's reduced activity, and in a lower sensitivity to a systemic shock.

11.3.2.3.3 Scenario 3: Trading Volume Peak in the PS (September 2009)

September 2009, the third scenario, is characteristic of a time when the volume traded in the CUD reached an all-time high. During this period the network continued to show an increase in distance, but with higher connectivity. The new properties of the network, along with the premise that failure-to-pay by a systemically important institution tends to be magnified during periods of high PS activity, should have exhibited a SP that is more vulnerable than in the two previously analyzed scenarios.

Nevertheless, the results of the simulation show that failure-to-pay by the selected institutions in this period affected only – on average – 22.6% of the institutions participating in the CUD. This is significantly lower than the 41.3% and 33.9% found in the TES market boom and stress scenarios, respectively (Figure 11.8).

Figure 11.8 **Attacks' impact on the institutions (Scenario 3)**



Note: Results truncated to 200% for practical purposes.
Source: authors' calculations.

This outcome is explained by a larger opening balance for CUD participants. In the boom and stress scenarios the average beginning-of-the-day funds in the CUD accounts corresponded to 36.4% and 24.2% of the September 2009's average.[27] In the peak trading volume scenario the additional opening balance provided the institutions and the PS with enhanced protection against the failure-to-pay of any institution, which resulted in fewer institutions being affected and more resilience of those being affected. It is worthwhile to stress that customary econometric models or NT alone would have disregarded this relevant feature of the system, which supports the choice of simulation models for capturing non-linearity and context-dependency.

11.3.2.3.4 Overall systemic risk assessment

Conveniently combining NT with the simulation approach makes it possible to identify and measure the importance of certain variables commonly overlooked by customary use of econometric models and NT. Such variables can mitigate or augment the systemic effect of an attack on TCTF institutions, and alter institutions' resilience and the capacity of the financial authorities to contain systemic risk. Network stability depends not only on the basic properties of the network (distance, diameter and connectivity), usually calculated through NT alone. It is also contingent on (i) network activity (eg number of participants and transactions, number and volume of payments); (ii) distribution of number of connections per node; (iii) initial conditions (eg institutions' opening balances in the CUD); (iv) the specialty of each business (eg managing third parties' portfolios); (v) the financial strategies used by participating institutions (eg the size of their portfolio of liquid securities); (vi) the regulation in place (eg being eligible for LLR liquidity or not); and (vii) the participants' behavior resulting from the arrival of adverse information (eg knowledge of a systemic event).

The results illustrate that external factors can overshadow the influence exerted by the network's intrinsic properties. Accordingly,

[27] Such a difference in the September 2009's opening liquidity is due to CIs maintaining larger deposits in the CUD during this period, which went from 5.95% of liabilities subject to reserve in February 2006 to 5.66% in June 2006 and 8.24% in September 2009. CIs are allowed to use their reserve requirements to meet their intra-day liquidity needs provided that their average effective reserve at the end of the bi-weekly period never drops below the required reserve.

the limited degree of network activity in the TES stress scenario proved to be a decisive variable in mitigating the impact of failure-to-pay by a systemic relevant institution, while the opening balance in the CUD during the peak trading scenario was the variable that did much to offset the increase in PoQ generated by the attacks and to make institutions more resilient.

This situation is evident in Table 11.6, which includes the results of the failures-to-pay for each selected scenario. There was more of an impact on the PS in the boom scenario (February 2006), when 41.3% of the CUD participants were affected, compared to 33.9% and 22.6% in the stress and peak volume scenarios, respectively. Such outcome may be explained by high counterparty exposure resulting from a boom period, along with low levels of opening balances in the CUD and the assumption of non-reaction from other participants.

In all the periods, BFs and MFs are the institutions affected the most by the attacks to central (TCTF) institutions. These were also the types of institutions where the variation in PoQ most often exceeded the TES portfolio and the OMO limit (ie they were the most impacted); this is, BFs and MFs are the less resilient institutions. BFs' resilience is hampered by the existing OMO limit, whilst MFs is hindered by the level of their TES portfolio. This is due mostly to the specific nature of their business, since BFs manage large own-account portfolios, whilst MF's portfolios are mainly third parties'. Therefore, when examining the behavior of the network in the different scenarios, it's evident that although contagion depends on factors external to the properties of the network, there are other explanatory factors such as the specific nature of the business of participating institutions and their regulations.

A look at the outcome of the attacks during the three selected periods, according to the type of institution (Table 11.7), shows the attack on the systemically important BFs affected – on average – 34.3% of the institutions, whilst the attack on the CBs affected 30.2%. It's important to emphasize that the TBTF concept would have missed this result, since the total assets (total investments) of the largest BF is about 16% (63%) of the CBs' average; this is also true for NT or simulation techniques based on balance sheet claims, since the balance sheet exposure of other institutions to BFs is non-large and collateralized.

Table 11.6 **Attacks' impact on the institutions – per scenario (vertical), per type of institution (horizontal)[a]**

Scenario	Criteria		Credit institutions (CI)			Non-credit institutions (NCI)			Aggregate	
			CB	CFC	CF	PFM	MF	BF		
TES boom (Feb 2006)	Affected institutions (mean)		55,9%	3,9%	62,5%	62,5%	46,0%	46,8%	41,3%	‡
	ΔPoQ / TES Portfolio*	Mean	7,5%	67,8%	11,1%	5410,8%	516,6%	141,6%	512,0%	‡
		Max.	48,4%	258,7%	21,8%	26605,7%	9752,3%	1497,5%	26605,7%	§
	ΔPoQ / OMO limit	Mean	10,0%	1,3%	15,8%	24,9%	137,4%	230,7%	109,7%	‡
		Max.	165,0%	2,9%	28,9%	78,7%	1734,9%	4186,7%	4186,7%	§
	ΔPoQ / TLF limit	Mean	19,2%	102,1%	37,2%	N/A[b]			61,6%	‡
		Max.	103,5%	399,3%	69,8%				399,3%	§
TES stress (June 2006)	Affected institutions (mean)		56,3%	2,5%	66,7%	25,0%	35,9%	40,6%	33,9%	‡
	ΔPoQ / TES Portfolio*	Mean	3,5%	200,5%	2,5%	203,1%	727,9%	2652,0%	1079,8%	‡
		Max.	21,5%	1579,2%	5,7%	396,3%	15916,4%	191641,1%	191641,1%	§
	ΔPoQ / OMO limit	Mean	4,9%	4,3%	2,8%	35,0%	92,9%	240,5%	103,1%	‡
		Max.	37,5%	17,1%	5,7%	35,0%	978,6%	3321,4%	3321,4%	§
	ΔPoQ / TLF limit	Mean	12,7%	26,9%	12,5%	N/A[b]			20,2%	‡
		Max.	103,5%	179,4%	30,4%				179,4%	§
CUD trading volume peak (Sept 2009)	Affected institutions (mean)		25,0%	2,3%	33,3%	50,0%	22,9%	22,9%	20,7%	‡
	ΔPoQ / TES Portfolio*	Mean	7,6%	6,0%	18,1%	118,8%	365652,2%	16,9%	85215,7%	‡
		Max.	26,4%	13,0%	24,0%	272,5%	4384820,5%	69,3%	4384820,5%	§
	ΔPoQ / OMO limit	Mean	10,6%	2,3%	47,9%	8,4%	31,1%	85,7%	37,9%	‡
		Max.	52,4%	4,9%	63,7%	26,1%	330,2%	286,4%	330,2%	§
	ΔPoQ / TLF limit	Mean	27,0%	5,3%	108,6%	N/A[b]			21,4%	‡
		Max.	139,2%	11,6%	144,3%				144,3%	§

[a] Institutions with TES portfolio or OMO limit equal to zero were discarded. Aggregate corresponds to the arithmetic sum of each type of institution (¤); to the number of institutions weighted average (‡); and to the maximum of all types of institutions (§). TES Portfolio* corresponds to the (1–α) market value of the TES Portfolio, where α is the BR's haircut. [b] TLF is reserved exclusively for Credit Institutions.

Source: authors' calculations.

Table 11.7 **Attacks' impact on the institutions – per attacked institution type (vertical), per type of institution (horizontal)[a]**

Attacked institution type	Criteria		Credit institutions (CI)			Non-credit institutions (NCI)			Aggregate	
			CB	CFC	CF	PFM	MF	BF		
CB	Affected institutions (mean)		45,8%	13,4%	58,3%	55,6%	36,1%	31,8%	30,2%	†
	ΔPoQ / TES Portfolio*	Mean	6,2%	5,5%	10,4%	1259,0%	190,3%	62,4%	142,1%	‡
		Max.	22,4%	12,0%	20,4%	6836,3%	2122,1%	623,1%	8239,9%	§
	ΔPoQ / OMO limit	Mean	8,5%	2,0%	22,8%	13,4%	77,5%	134,6%	64,1%	‡
		Max.	57,2%	4,6%	54,2%	24,4%	547,9%	770,0%	770,0%	§
	ΔPoQ / TLF limit	Mean	20,2%	4,7%	53,6%	N/A[b]			13,8%	‡
		Max.	121,3%	10,7%	122,8%				123,9%	§
BF	Affected institutions (mean)		47,9%	2,5%	54,2%	48,6%	43,8%	49,0%	34,3%	†
	ΔPoQ / TES Portfolio*	Mean	6,2%	177,4%	9,6%	2002,5%	244407,5%	2704,3%	59089,1%	‡
		Max.	33,1%	918,9%	17,6%	19941,1%	2197286,4%	96016,7%	2288230,8%	§
	ΔPoQ / OMO limit	Mean	8,5%	3,0%	16,2%	26,2%	96,7%	307,4%	124,5%	‡
		Max.	100,1%	8,6%	32,2%	68,0%	1356,8%	2478,3%	2582,7%	§
	ΔPoQ / TLF limit	Mean	19,1%	84,9%	40,2%	N/A[b]			54,8%	‡
		Max.	98,4%	289,4%	75,0%				289,4%	§

[a] Institutions with TES portfolio or OMO limit equal to zero were discarded. Aggregate corresponds to the arithmetic sum of each type of institution (¤); to the number of institutions weighted average (‡); and to the maximum of all types of institutions (§). TES Portfolio* corresponds to the (1–α) market value of the TES Portfolio, where α is the BR's haircut. [b] TLF is reserved exclusively for Credit Institutions.
Source: authors' calculations.

CFs were the institutions most affected (ie share of CFs which experimented an increase in their PoQ), followed by the BFs and the PFMs. As in the analysis by periods, the BFs and the MFs are the institutions whose liquidity is affected the most. The attack on the systemically relevant BFs resulted in a situation where non-systemically relevant BFs exhibited, on average, a variation in PoQ higher than their TES portfolio and than their OMO limit. The situation with the CFCs and PFMs was similar.

Attacks on CBs had a similar impact on the MFs and PFMs. It is to note that the PFMs are less active in the CUD, and do not routinely resort to transitory expansion operations for liquidity, given the volume and liquidity of the resources they manage and the –long and immovable- maturity of the portfolios they manage. A significant share of the institutions that experienced more of an increase in their PoQ were found to be resilient; this is, they have alternatives for solving the liquidity strains generated by the attacks, be it through the available TES portfolio or through the OMO and TLF facilities (eg CFCs, CFs and CBs).

11.4 Concluding remarks

Systemic risk is a negative externality. Financial market's participants have clear incentives to manage their own risk (eg credit, market, legal, operational, etc.), but no incentives exist for them to account the effects of their actions on other institutions or the system as a whole; this is, each individual institution is clearly motivated to prevent its own collapse but not necessarily the collapse of the system as a whole (Trichet, 2009). As a social consequence of individual behavior, systemic risk has to be addressed within a comprehensive approach, capable of capturing institutions' contribution to systemic risk.

The most recent episode of market turmoil exposed the limitations resulting from using micro-prudential approaches (eg TBTF) to identifying and assessing institutions' contribution to systemic risk when applied to payments and financial systems, which are characterized by high levels of complexity, cross-dependency, context-dependency and non-linearity. Such limitations are not new, but have been increasingly important overtime due to the escalating homogeneity and opaqueness of financial markets, which has resulted in what the authors consider as an increasingly systemic-crisis-prone financial system.

It is rather clear that a qualitative leap towards a more broad and comprehensive analysis of financial markets and payments systems is the first step to effectively identifying and assessing systemic risk. Accordingly, this document proposes an approach consisting of applying a convenient mixture of NT and simulation techniques, where the former allowed for identifying those institutions that can be regarded as central for the system, and the latter allowed for assessing and analyzing the resilience of not-attacked institutions and of the system as a whole. Afterwards, based on quantitative assessment of individual liquidity requirements, central bank's means for containing systemic risk via its liquidity facilities are appraised and analyzed.

Results of this approach when applied to the Colombian large-value payments system (CUD) yield three main remarks. First, results confirm that customary micro-prudential approaches (ie institution centric) are insufficient for identifying and assessing sources of systemic risk. Second, results draw attention to an ongoing debate on the improvement of the financial systems' resilience through an adequate liquidity provision framework. Third, because this is a preliminary approach, some challenges for further research are still pending. This remarks will be discussed next.

11.4.1 The importance of macro-prudential approaches

Results converge to recent literature on systemic risk: although the size of the institution influences the systemic importance of the participating institutions (ie CBs), market activity and connectedness within the network play a key role in defining systemic relevance (ie BFs). This is, the connections between financial institutions are as important as the institutions themselves.

Results also confirmed the importance of developing macro-prudential approaches. Unlike traditional micro-prudential approaches (eg TBTF), the Colombian case displayed that BFs are systemic risk sources as important as CBs. Despite the role of BFs was previously believed as systemically important within Colombia's financial system, this document provides an innovative approach that allows for quantitatively assessing systemic importance at a disaggregate level (by institution). Furthermore, this results highlight the importance of the 'shadow banking system', where too-connected institutions, regardless of their size or the value of their claims held by other participants, may endanger the safety of the payments system and financial stability.

It is important to recognize that reaching a true macro-prudential approach to systemic risk requires a coordinated supervision and regulation of the financial system. Financial authorities should work together in order to establish a comprehensive regulatory framework that aims to the efficient and safe functioning of the financial system, whereas supervision should be capable of effectively tracking individual and collective behavior for preserving such framework, where assessment and enforcement tools are key for this task. Such coordination may require designing clear institutional arrangements.

The approach also allowed for an inclusive characterization of payments systems. Their stability depends not only on the basic properties of the network (distance, diameter and connectivity). It is also contingent on (i) network activity (eg number of participants and transactions, number and volume of payments); (ii) distribution of number of connections per node; (iii) initial conditions (eg institutions' opening balances in the CUD); (iv) the specialty of each business (eg managing third party's portfolios); (v) the financial strategies used by participating institutions (eg the size of their portfolio of liquid securities); (vi) the regulation in place (eg being eligible for LLR liquidity or not); and (vii) the participants' behavior resulting from the arrival of adverse information (eg knowledge of a systemic event).

Abbreviations

BF	Brokerage Firm (please refer to Table 11.1)
BR	Banco de la República (Colombia's central bank)
CB	Commercial Bank (please refer to Table 11.1)
CF	Financial Corporation (please refer to Table 11.1)
CFC	Commercial Financial Corporation (please refer to Table 11.1)
CHAPS	Clearing House Automated Payment System
CI	Credit Intermediaries (banking institutions, please refer to Table 11.1)
CPSS	Committee on Payment and Settlement Systems
CUD	Colombia's large-value payment system
DCV	BR's depositary for clearing and delivering of TES
Deceval	Private depositary for clearing and delivering stocks and TES
ECB	European Central Bank
FIFO	First-In-First-Out
IMF	International Monetary Fund
LLR	Lender-of-Last-Resort
LTCM	Long-Term Capital Management
MF	Mutual Fund (please refer to Table 11.1)
NCI	Non-credit Intermediaries (non-banking institutions, please refer to Table 11.1)
NT	Network Topology
OMO	Open Market Operations
PFM	Pension Fund Manager (please refer to Table 11.1)
PoQ	Payments on Queue
PS	Payments System
QRA	Queue Resolution Algorithm
RTGS	Real-Time Gross Settlement system
SEN	BR's Electronic Negotiation System (TES only)
TBTF	Too-big-to-fail
TCTF	Too-connected-to-fail
TES	Colombia's central government local bond
TLF	Transitory Liquidity Facility

References[1]

Acharya, V V – Philippon, T – Richardson, M – Roubini, N (2009) **The Financial Crisis of 2007–2009: Causes and Remedies.** Restoring Financial Stability (Eds. Acharya, V.V. and Richardson, M.), Wiley Finance.

Arjani, N – Engert, W (2007) **The Large-Value Payments System: insights from selected Bank of Canada research.** Bank of Canada Review, Bank of Canada, Spring.

Battiston, S – Delli Gatti, D – Gallegati, M – Greenwald, B C – Stiglitz, J E (2009) **Liasons Dangereuses: increasing connectivity, risk sharing and systemic risk.** NBER Working Papers, Working Paper 15611, January.

Bech, M – Garrat, R (2006) **Illiquidity in the interbank payment system following Wide-Scale Disruptions.** Federal Reserve Bank of New York Staff Report, No. 239, March.

Becher, C – Millard, S – Soramäki, K (2008) **The network topology of CHAPS Sterling.** Working Paper, No. 355, Bank of England, November.

Boss, M – Elsinger, H – Summer, M – Thurner, S (2004) **An empirical analysis of the network structure of the Austrian Interbank market.** Financial Stability Report, No. 7, ONB.

Brunnermeier, M – Crocket, A – Goodhart, C – Persaud, A D – Shin, H (2009) **The Fundamental Principles of Financial Regulation.** International Center for Monetary and Banking Studies, Centre for Economic Policy Research (CEPR), January.

Bullard, J – Neely, C J – Wheelock, D C (2009) **Systemic Risk and the Financial Crisis: A Primer.** Federal Reserve Bank of St. Louis Review, Federal Reserve Bank of St. Louis, September/October 2009.

[1] Authors' preliminary versions of published documents (*) are available online (http://www.banrep.gov.co/publicaciones/pub_borra.htm).

Cepeda, F (2008)* **La topología de redes como herramienta de seguimiento en el Sistema de Pagos de Alto Valor en Colombia.** Borradores de Economía, No. 513, Banco de la República.

Chan-Lau, J (2010a) **Financial Crisis, Interconnectedness and Regulatory Capital.** Lombard Street, Vol. 2, No. 1.

Chan-Lau, J (2010b) **Balance Sheet Network Analysis of Too-connected-to-fail Risk in Global and Domestic Banking Systems.** IMF Working Paper, WP/10/107, International Monetary Fund, April

Clark, J (2010) **Too-networked-to-fail.** Risk <online>, June 7th.

Clement, P (2005) **The term 'macroprudential': origins and evolution.** BIS Quarterly Review, Bank for International Settlements, March.

Committee on Payment and Settlement Systems (CPSS) (2005) **Central bank oversight of payment and settlement systems.** Bank for International Settlements, May.

Committee on Payment and Settlement Systems (CPSS) (2001) **Core Principles for Systemically Important Payment Systems.** Bank for International Settlements, January.

Dowd, K (2002) **Measuring Market Risk.** John Wiley & Sons.

Elton, C S (1958) **The Ecology of invasions by animals and plants.** Methuen.

European Central Bank (ECB) (2010) Financial Stability Review, June.

Gai, P – Kapadia, S (2010) **Contagion in financial networks.** Working Paper No. 383, Bank of England, March.

Haldane, A G (2009) **Rethinking the financial network.** Speech delivered at the Financial Student Association (Amsterdam, Netherlands), April.

Ianoka, 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 Paper Series, No. 04-E-04, April.

International Monetary Fund (IMF) (2009a) **Initial lessons of the crisis.** February 6th.

International Monetary Fund (IMF), Bank for International Settlements (BIS); Financial Stability Board (FSB) (2009b) **Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations.** October.

Jurgilas, M – Martin, C (2010) **Liquidity-saving mechanisms in collateral-based RTGS payment systems.** Bank of England Working papers No. 389, March.

Kambhu, J – Weidman, S – Krishnan, N (2007) **New directions for understanding systemic risk.** The National Academies Press.

Kodres, L (2009) **The Road to Re-Regulation: Repainting the Center Line and Erecting more Guardrails.** The World Bank and Banco de España International Conference: 'Reforming Financial Regulation and Supervision: Going Back to Basics', Madrid, June 15.

Krugman, P (2009) **The return of depression economics and the crisis of 2008.** W.W. Norton.

Landau, J-P (2009) **Complexity and the financial crisis.** Introductory remarks at the Conference on The Macroeconomy and Financial Systems in Normal Times and in Times of Stress, Banque de France and Bundesbank, June 8th.

Leinonen, H – Soramäki, K (2004) **Simulation: a powerful research tool in payment and settlement systems.** Journal of Financial Transformation, Vol. 12.

Lublóy, Á (2006) **Topology of the Hungarian large-value transfer system.** Magyar Nemzeti Bank Occasional Paper, No. 57

Machado, C L – León, C – Sarmiento, M – Cepeda, F – Chipatecua, O – Cely, J (2010)* **Riesgo sistémico y estabilidad del sistema de pagos de alto valor en Colombia: análisis bajo topología de redes y simulación de pagos.** Borradores de Economía, No. 627, Banco de la República.

Manning, M – Nier, E – Schanz, J (2009) **The economics of large-value payments and settlement: Theory and policy issues for central banks.** Oxford University Press.

Meltzer, A (1986) **Financial Failures and Financial Policies.** Deregulating Financial Services: Public Policy in Flux (Eds. Kaufman, G G and Kormendi, R C), Ballinger.

Pozen, R (2010) **Too big to save? How to fix the U.S. financial system.** John Wiley & Sons.

Reveiz, A – León, C (2008)* **Índice representativo del mercado de deuda pública interna: IDXTES.** Borradores de Economía, No. 488, Banco de la República.

Saunders, A – Smith, R C – Walter, I (2009) **Enhanced Regulation of Large, Complex Financial Institutions.** Restoring Financial Stability (Eds. Acharya, V V and Richardson, M), Wiley Finance.

Schmitz, S – Puhr C (2009) **Structure and stability in payments network – A panel data analysys of ARTIS simulations.** (May 7). Available at SSRN: http://ssrn.com/abstract=1400883 or http://dx.doi.org/10.2139/ssrn.1400883.

Soramäki, K – Bech, M – Arnold, J – Glass, R – Beyeler, W (2006) **The topology of interbbank payments flow.** Federal Reserve Bank of New York Staff Report, No. 243, March.

Soramäki, K – Beyeler, W – Bech, M – Glass, R (2007) **New approaches for payment system simulation research.** Simulation studies of liquidity needs, risks and efficiency in payment networks (Ed. Leinonen, H), Scientific monographs E:39, Bank of Finland.

Sornette, D (2003) **Why Stock Markets Crash.** Princeton.

Trichet, J-C (2009) **Systemic risk.** Text of the Clare Distinguished Lecture in Economics and Public Policy, University of Cambridge, Cambridge, December 10th.

Tumpel-Gurell, G (2009) **Introductory remarks, Recent advances in modeling systemic risk using network analysis** (workshop), European Central Bank, Frankfurt am Main, October 5th.

World Bank (2009) **The evolution of Real-Time Gross Settlement.** Financial Infrastructure Series, February.

Zhou, C (2009) **Are banks too big to fail? Measuring systemic importance of financial institutions.** De Nederlandsche Bank.

Chapter 12

Operational risk in ReGIS – a systemically important payment system

Horațiu Lovin – Andra Pineta**

* The authors are grateful to Bank of Finland for providing technical support and expertise in order to implement in Romania the BoF-PSS2 simulator, an instrument developed to better understand risks faced by payment and settlement systems.

12 Operational risk in ReGIS – a systemically important payment system

Abstract

The study aims to quantify the impact of an operational incident on the ReGIS payment system and to assess its ability to absorb a liquidity shock. Operational risk tends to increase in importance, relative to other risks (eg legal risk, economic risk), as a payment system expands. The related back-up systems and contingency plans ensure that the probability of an operational incident with significant impact on the smooth functioning of the payment system is small, but close monitoring is crucial because of the unpredictable nature of such incidents. The data for this study cover participant's transactions and liquidity resources in September–December 2008. The results reveal a contraction of ReGIS liquidity since October 2008 but also that the payment system has a strong capacity to absorb a medium-intensity liquidity shock due to an operational incident. Only one participant is systemically important, but its being hit by a severe operational incident could trigger a liquidity shock in ReGIS, with significant negative effects such as an inability to settle some payment orders, a rise in money market interest rates or a decrease in interbank landing.

12.1 Introduction

Smooth operation of payment systems is of particular importance for central banks. Current research is focusing on the development and modernization of payment systems, as well as on identifying potential risks and inefficiencies due to the architecture and/or operating modes of payment systems. The most recent work in the field has been in the testing of payment system performance in terms of risk control and resistance to shocks.

Romania's national payment system ReGIS is crucial for financial stability, because it ensures settlements of central-bank monetary policy operations, interbank payments, net positions of all payment and clearing systems, and fund transfers related to financial instrument (securities) transactions among the securities settlement systems.

ReGIS[1] is an RTGS system designed to ensure the processing and real-time gross settlement of large-value (over RON 50,000) and urgent payment instructions sent by participants, as well as instructions related to ancillary systems.[2]

As regards the control of operational risk, the national payment system has been designed in accord with international security standards so as to maintain a high level of operational resilience. Thus, the architecture of the electronic payment system includes a contingency processing center (secondary site), which enables the resumption of data processing achieved by the automatic copying of information from the main operating center (primary site). Although operational risk (from a technical and technological viewpoint) is greatly diminished, human errors are still likely to occur. In order to prevent these, working procedures have been improved by clearly defining the activities and operations of system participants and the system operator. Currently, there are operating procedures, contingency plans, back-up procedures in case of disaster, and plans to ensure the business continuity of ReGIS.

Smooth functioning of payment systems also involves the availability of resources such as buildings, experts, special IT equipment, electricity etc. All these resources are exposed to operational risks (idiosyncratic or not) that can cause business disruptions. For their own protection, participants have the means for archiving data, generating electricity in an emergency, and outsourcing certain activities. Due to the diversity and low probability of occurrence of extreme operational incidents, the costs of protective measures may exceed the benefits.

The purpose of this analysis is to assess the impact of such extreme events on ReGIS, using a software application developed by the Bank of Finland, the simulator for payment and settlement systems (version BoF-PSS2). Simulation techniques enable us to build a realistic operating environment, which can be used for observing and testing scenarios with a view to assessing the system's capacity to absorb liquidity shocks. The BoF-PSS2 simulator, one of the most highly regarded tools for analysing payment systems in terms of liquidity and contagion risk, is currently used by central banks in over 60 countries. Soramaki and Leinonen (2003) provide a comprehensive description of the simulator.

[1] For a comprehensive description of ReGIS see http://www.bnr.ro/ReGIS-3305.aspx.
[2] Auxiliary systems are SENT, SaFIR, RoClear, PCH, MasterCard, Visa.

The scenarios for operational incidents were aimed at analyzing the ability of the payment system to absorb liquidity shocks. Due to the lack of major incidents (so far) in ReGIS, the scenarios have not generated occurrence probabilities. Participant's behavior, as well as the general features, such as incentives to protect liquidity reserves, to reduce borrowing costs, and to isolate participants with liquidity shortages has been assumed to be uniform.

Koponen and Soramaki (1998) first used the simulator to quantify the impacts of operational incidents on the Finnish payment system and proposed indicators of liquidity tensions in the payment system. Bedford, Millard and Yang (2004) studied the systemic effects of an operational incident which affects one participant and observed the other participants' reactions. They used data from the CHAPS UK payment system and found that only simultaneous disruption of three major banks could cause significant systemic losses. Glaser and Haene (2008) adapted the scenario assumptions to the particularities of the Swiss payment system and expanded the time reaction of participants not directly affected by the incident and assessed the daily moment when an operational incident would cause the maximum systemic loss. Lubloy and Tana (2007) identified systemically important participants in Hungary, quantitatively assessed the system's ability to cope with a liquidity shock, and computed the amount of additional liquidity needed to settle all the transactions when payment system faces a liquidity shock.

This study assesses the impact on ReGIS of the global financial crisis triggered by the collapse of Lehman Brothers, using matrix analysis and volatility computations. At the same time, it tests the payment system's resilience via a scenario approach consistent with the patterns / concepts / findings / studies / methods / knowledge of the recent and relevant the Simulator research literature.

The data for this study cover, for September to December 2008, the participants' transactions and liquidity resources. Section 12.2 presents a general description of the ReGIS payment system during September – December 2008, including its size, liquidity and concentration indicators, and issues related to participants' behavior. In Section 12.3 we analyze the period 13–31October 2008, during which liquidity tightened in the banking system, and Sections 12.4 and 12.5 assess the payment system's capacity to absorb medium-sized liquidity shocks (Section 12.4) and severe liquidity shocks (Section 12.5), all caused by operational incidents. Conclusions are presented in Section 12.6.

12.2 ReGIS payment system during 1 September – 31 December 2008

International financial imbalances have triggered a global liquidity shortage and an increase in risk aversion throughout the financial system. Increasing perceived counterparty risk in the financial system can acerbate a liquidity shortage and cause market failures via a build-up of excess liquidity reserves and a freezing up of interbank lending. The domestic financial system by 2008 was benefiting from a liquidity surplus caused by inflows from non-resident investors, which allowed smooth payment system operation. A highly integrated European, and more importantly global, financial system combined the known benefits with new vulnerabilities, namely significant exposures to external liquidity shocks.

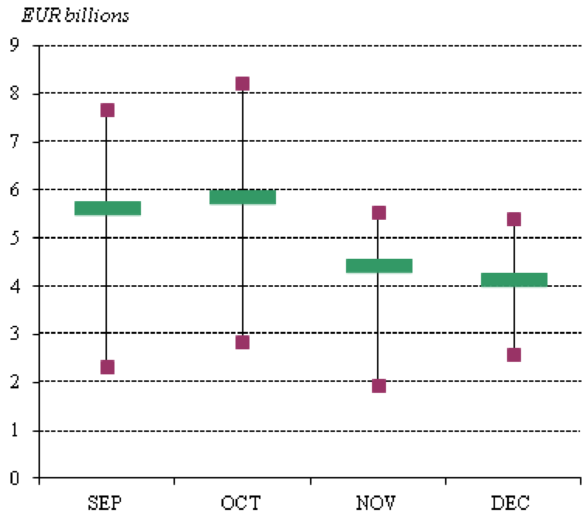
Payment values were high in October 2008, when significant tensions in the money market required an important intervention of the central bank, which absorbed liquidity from participants with excess reserves and granted loans to participants with liquidity shortages. The volatility of the number of transactions was higher in December 2008, which might be usual for the winter holiday season. (Chart 12.1 and 12.2).

Observing the daily transaction patterns, the behavioral model can assess participants' preferences as to initiating payments. The low value payments are settled at the beginning of the day, while higher value payments are settled at the end of the day (Chart 12.3). It is possible for participants to use this strategy in order to reduce liquidity risk and the cost of borrowing in the interbank market. Payments are settled in the first part of the day using resources available at the beginning of the day, but for medium and large value payments, the participants take into account payments received from other participants in order to minimize borrowing costs. They also need to strike a balance between liquidity management costs and the quality of services provided to clients, because of the reputational risk implied by payment delays.

The top four participants in the payment system, in terms of value of payments submitted, have a 50 percent market share, and the top 12 participants accumulated about 90 percent of total payments (Chart 12.4). The hierarchy in the payment system differs from that of the banking system because there are banks (subsidiaries of major international financial groups) that are highly active in the payments system do not have large asset holdings.

Chart 12.1

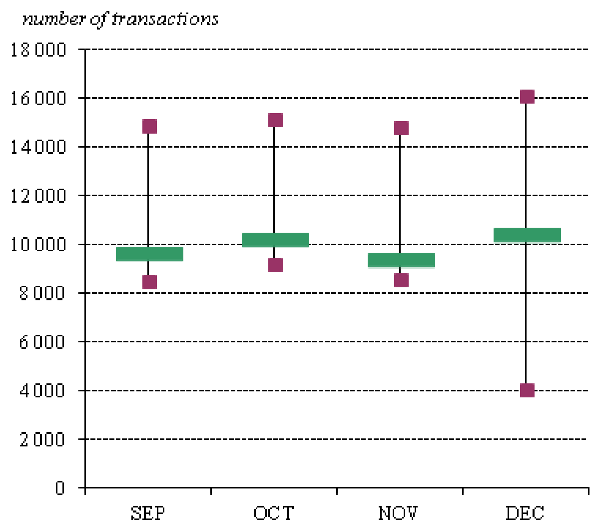
Daily payment values settled in ReGIS, September – December 2008



Note: The chart contains maximum, minimum and median values for the indicator. This note applies also for charts 2, 9–12, 14–16.
Source: National Bank of Romania

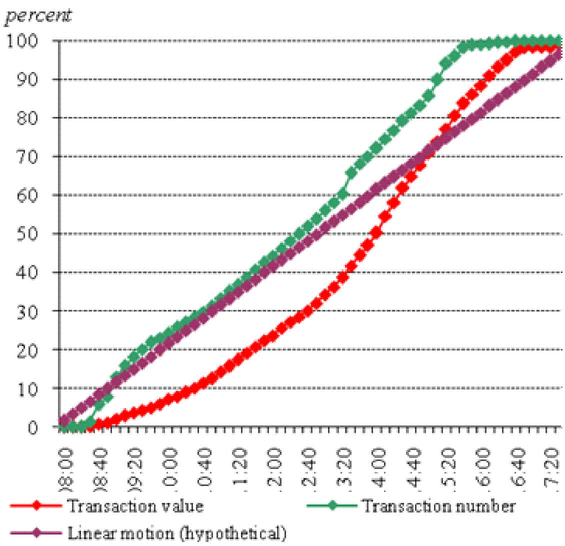
Chart 12.2

Daily numbers of payments settled in ReGIS, September – December 2008



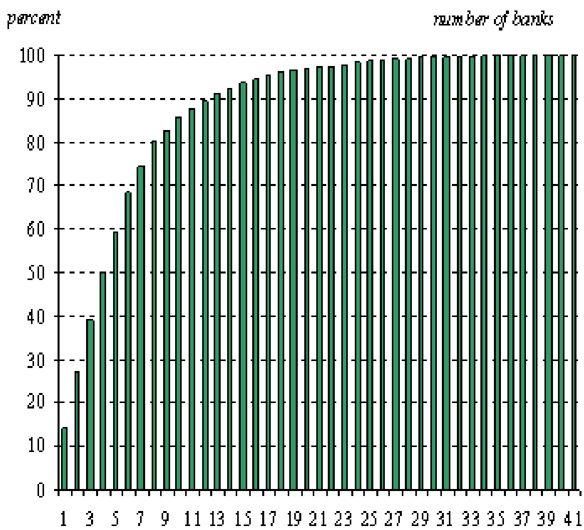
Source: National Bank of Romania

Chart 12.3 **Cumulated value and number of payments during the day**



Source: National Bank of Romania

Chart 12.4 **Concentration of ReGIS payment system (excl. State Treasury and central bank)**

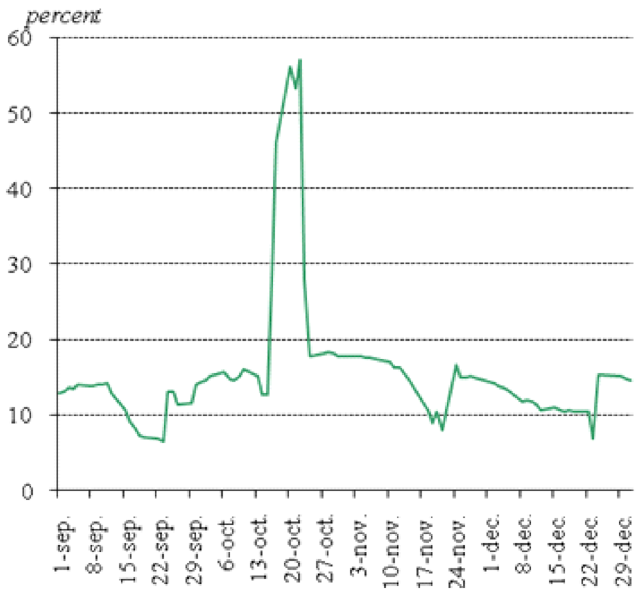


Source: National Bank of Romania

12.3 Did the credit risk perceived by participants increase when liquidity tightened during the period 13–31 October 2008?

The analysis reveals a liquidity tightening in the payment system in 13–31 October 2008. Interest rates rose rapidly (Chart 12.5), and the operating hours for ReGIS were extended at the request of SaFIR[3] for settling transactions related to credit facilities granted to participants by the central bank (Table 12.1). Regarding open market operations, the central bank faced a period of transition from net debtor to net creditor of the banking system. The increasing government deficit and diminution of inflows from parent banks absorbed the excess liquidity.

Chart 12.5 **ROBOR-ON interest rate, September – December 2008**



Source: National Bank of Romania

[3] SaFIR is the central depository and settlement system for government securities, owned and operated by National Bank of Romania. SaFIR provides custody services for government securities and certificates of deposit issued by National Bank of Romania, and also settles the transactions with these financial instruments. In addition, SaFIR manages collateral to settle money market operations, including transactions with the central bank.

Table 12.1 **One-off extensions of ReGIS transitioning hours**

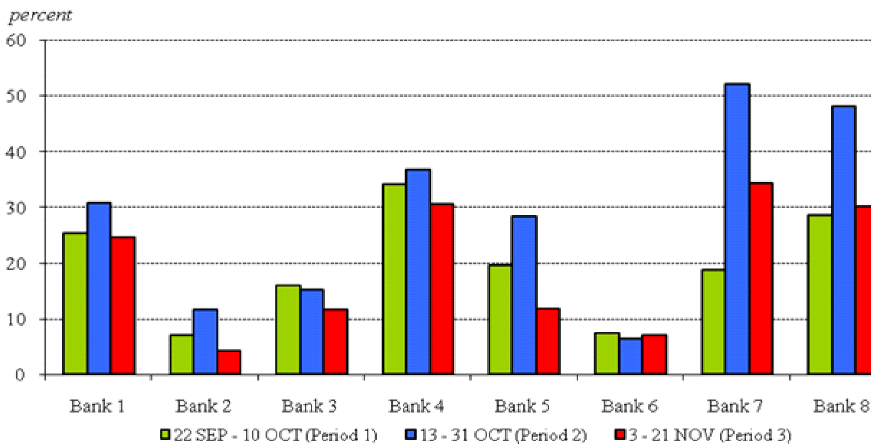
Date	Schedule extension
16 October 2008	5 min
17 October 2008	1 h and 25 min
20 October 2008	20 min
23 October 2008	10 min
24 October 2008	5 min
31 October 2008	50 min

Source: National Bank of Romania

At the international level, a liquidity shortage and bank losses increased credit risk for the financial institutions and the side effect has been a more severe liquidity contraction. The spiral thus created destabilized the global banking sector, which had a negative impact on emerging markets, including the Romanian financial market.

In order to test the increasing pressure on liquidity resources and perceived counterparty credit risk for payment system participants between 13 and 31 October 2008, we consider three observation periods: (1) 22 September – 10 October; (2) 13 – 31October and (3) 3 – 21 November. Periods 1 and 3 are not characterized by liquidity imbalances in the payment system. We measure intraday volatility of account balances for the first eight participants (credit institutions) in the payment system (accounting for about 80% of total transactions) to see whether liquidity tensions occurred during period 2. In addition, we analyze bilateral transactions among the first eight participants for all three periods in order to determine whether those transactions increased in period 2, affecting the smaller participants. In a highly uncertain environment, participants with lower volumes of transactions could be perceived as riskier, which would reduce their liquidity inflows.

Chart 12.6 **Account balances volatility for the first 8 participants in ReGIS, 22 September – 21 November 2008**



Source: National Bank of Romania

The account balances volatility during the day is higher in period 2 than in periods 1 and 3 for the most important participants in the payment system (Chart 12.6). The increased generalized volatility (6 of 8 participants) leads to the conclusion that liquidity declined, but only for a short period of time. Liquidity injections by the central bank were quickly felt in the payment system as they reduced pressure on resources. Noteworthy is that after that period, the central bank become a net creditor of the banking system, in terms of open market operations, after several years characterized by excess liquidity.

Table 12.2 **Relative bilateral transactions for the top 8 ReGIS participants, sorted in the first column, 22 September - 21 November 2008**

	Bank 1	**Bank 2**	**Bank 3**	**Bank 4**	**Bank 5**	**Bank 6**	**Bank 7**	**Bank 8**
Bank 1	–					Red	Green	
Bank 2	Green	–	Red			Red		
Bank 3		Red	–	Green			Red	Red
Bank 4	Green	Green	Green	–	Red	Red	Red	
Bank 5	Green	Green			–	Red		
Bank 6	Green		Green		Red	–	Green	Red
Bank 7	Green		Red				–	Red
Bank 8		Green	Red			Red		–

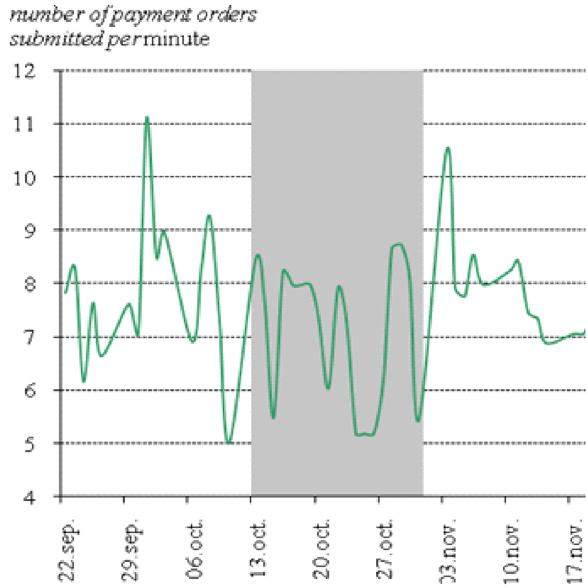
Note: Green indicates higher values of bilateral transactions in period 2 (13 – 31 October 2008) than in period 1 (22 September – 10 October 2008) and period 3 (3 – 21 November 2008); red indicates lower values of bilateral transactions in period 2 than in periods 1 and 3.
Source: National Bank of Romania

Bilateral transactions between the top eight participants in the payment system have not increased, to the detriment of the smaller participants (Table 12.2). Marginalization of smaller participants, more vulnerable in periods with high uncertainty, may signal an increase in perceived counterparty risk in the payment system due to information asymmetry.

Observing the account balance volatility for the top 8 participants in the payment system and their bilateral transactions, we find that in 13 – 30 October 2008, which was characterized by pronounced instability in the global financial system, there was a liquidity shortage, but no increase in perceived counterparty risk in the ReGIS payment system.

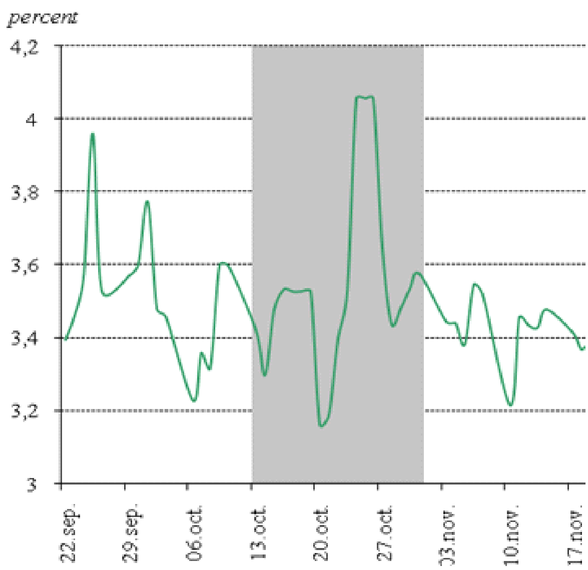
To support our conclusions, we computed the payments frequency (payments per minute) and frequency volatility (Charts 12.7 and 12.8). The data do not indicate a significant change in participants' behavior in 13 – 31 October 2008. The results should be viewed with caution because many factors can lead to random changes in payments frequency, with no connections to available liquidity.

Chart 12.7 **Payment orders frequency, 22 September – 21 November 2008**



Source: National Bank of Romania

Chart 12.8 **The volatility of Payment orders frequency, 22 September – 21 November 2008**



Source: National Bank of Romania

Further, we stressed the ReGIS payment system via two scenarios (medium and severe) for testing the system's ability to absorb liquidity shocks due to operational incidents. The period 13 – 31 October 2008 was analyzed separately for these two scenarios, to obtain important information about liquidity flows in the payment system during periods of liquidity stress.

12.4 Scenario 1

Based on the available liquidity at the beginning of the day for each participant in the payment system (credit institutions and State Treasury) and the payment orders entered into the system, the simulator replicated the settlements. The simulator applies the FIFO method (first in first out) to settle payments, and at the end of the day all outstanding payments are automatically settled, so that payment orders cannot be postponed to the next day.

In the scenario, payment-system replication is based on the following assumptions:

a) Participants can settle payments only with resources available in their accounts at the start of the day and payments received from other participants during the day
b) Large value payments cannot be split; a payment is are settled only if the payer has the entire amount in its account
c) The central bank is not exposed to liquidity risk; it can issue currency.

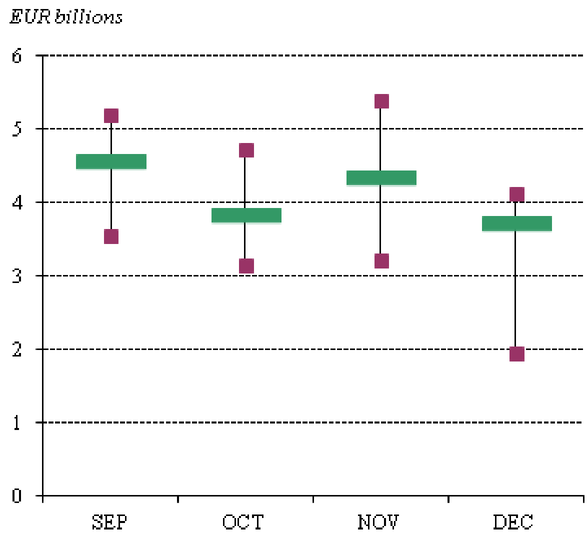
The scenario objective is to determine whether participants in the ReGIS had sufficient liquidity and to gauge the extent they used additional resources, as well as to examine the impact of a liquidity deficit on system functioning. Since the data sample covers the beginning of the crisis (October 2008), the payment system was considered to be already stressed, so that we look only at system functioning for evidence of payment-system resilience to liquidity shocks.

Liquidity used is the amount of financial resources effectively used by participants to settle the submitted payments. This amount is smaller than the total value of transactions because liquidity recycling enables multiple transaction settlements with the same money. The recycling ability of financial resources is an indicator of the efficiency of the payment system from the perspective of reducing both borrowing cost and the opportunity cost of holding cash, but a high degree of resource recycling may cause vulnerability to a sudden drying up of liquidity. Participants have rational incentives to reduce costs, which can also serve the interests of clients by decreasing their transaction costs. But a cautious balancing of risks and benefits is required in order to prevent taking excessive liquidity risk.

In ReGIS payment system, the degree of resources recycling was low during the analyzed period, due to excess liquidity and the high level of the reserve accounted by the central bank, but since October 2008 participants must deal with a less liquidity environment. The total amount of liquidity used in payment settlements and the liquidity usage indicator has been volatile during the period September – December 2008 (Chart 12.9 and 12.10) as a result of international financial turmoil, frequent liquidity injections provided by central bank and increasing budget deficit. We consider that liquidity usage indicator will stabilize in medium term, but at a higher level compared with the one before October 2008. We believe that this development of the indicator will have positive effects on the domestic payments because it will impose a more rigorous management of financial resources.

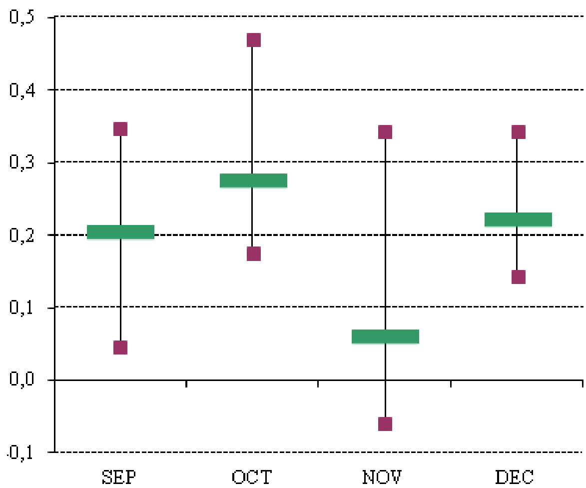
To quantify the impact of scenario assumptions, we use indicators like total queues size (maximum daily values and 10 minutes frequency values throughout the day), the number of participants with temporary liquidity shortages and minimum balances of the participant accounts.

Chart 12.9 **Daily liquidity used, September – December 2008**



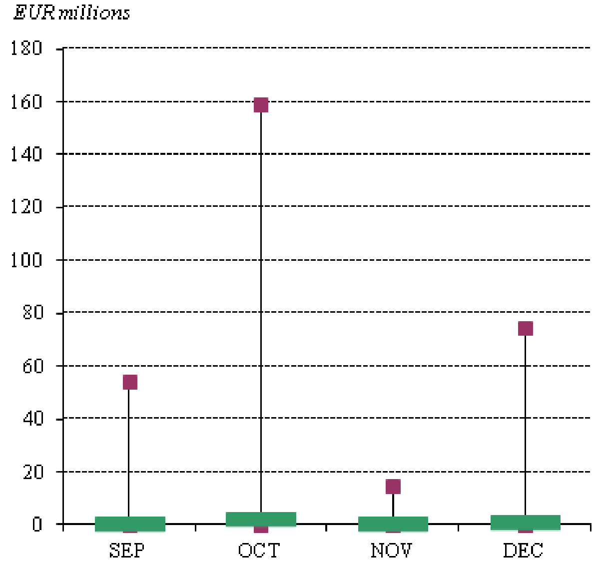
Source: National Bank of Romania

Chart 12.10 **Liquidity usage indicator, September – December 2008**



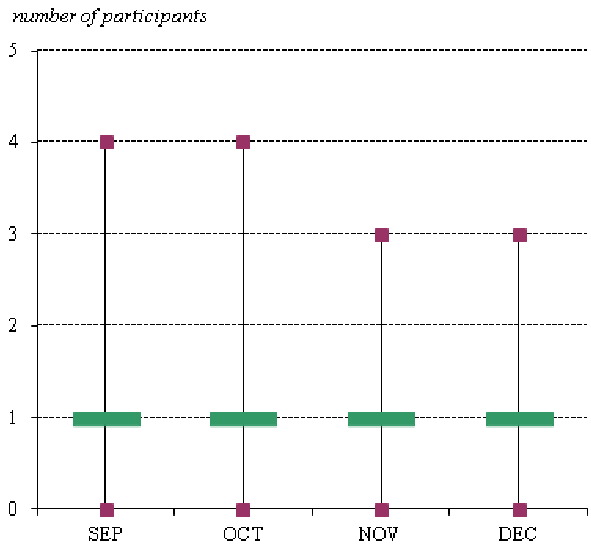
Note: Liquidity usage indicator is computed by dividing liquidity used at beginning of the day account balance
Source: National Bank of Romania

Chart 12.11 **Maximum daily queues, September – December 2008**



Source: National Bank of Romania

Chart 12.12 **Maximum daily participants with queued payment orders, September – December 2008**



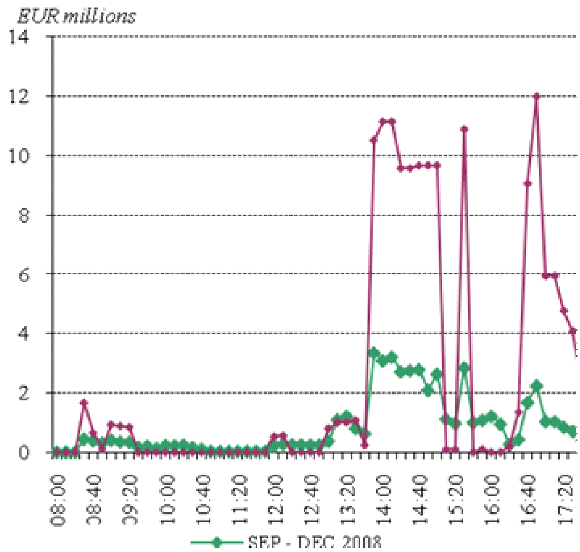
Source: National Bank of Romania

Liquidity tightening in October 2008 and significant injections by the central bank in October-November 2008 are captured by the scenario results. Queues increased significantly in October and December 2008 (Chart 12.11), but only a few participants were directly affected (Chart 12.12).

Queues increase in the second part of the day, when most of the large value payment orders are submitted, which increases the pressure on financial resources. The payment system is not affected by an operational incident if it takes place in the first half of the day, but tensions can arise if it occurs in the second half of the day. The impact was more pronounced in 13 – 31 October 2008, when the payment system had been already faced liquidity tightening. By the end of the day, the payment system fully absorbed the tension induced by the scenario assumptions (queue size reached zero), which is evidence of strong resistance to the liquidity shock (Chart 12.13).

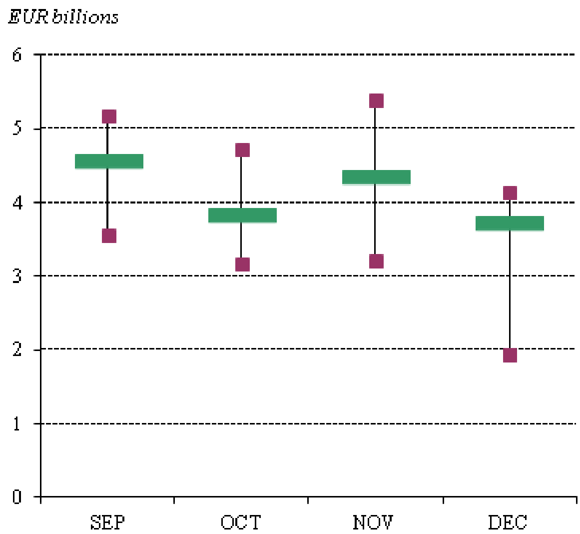
The evolution of the minimum balance on participants' accounts throughout the day reflects a liquidity shortage in October 2008 and a seasonal effect in December 2008, due to winter holidays (Chart 12.14).

Chart 12.13 **Shock transmission during the day (data available every 10 minutes)**



Source: National Bank of Romania

Chart 12.14 **Minimum daily account balance, September – December 2008**



Source: National Bank of Romania

12.5 Scenario 2

This scenario produces a severe liquidity shock, under the following assumptions:

a) An operational incident disrupts the IT system of the three largest participants in ReGIS (by total payments submitted, excl. central bank and State Treasury), which are then unable to submit payments into the system; the incident impacts only one participant each time, therefore the scenario was run three times

b) The other participants do not observe the operational incident and continue to make payments to the affected participant but are unable to receive payments from it

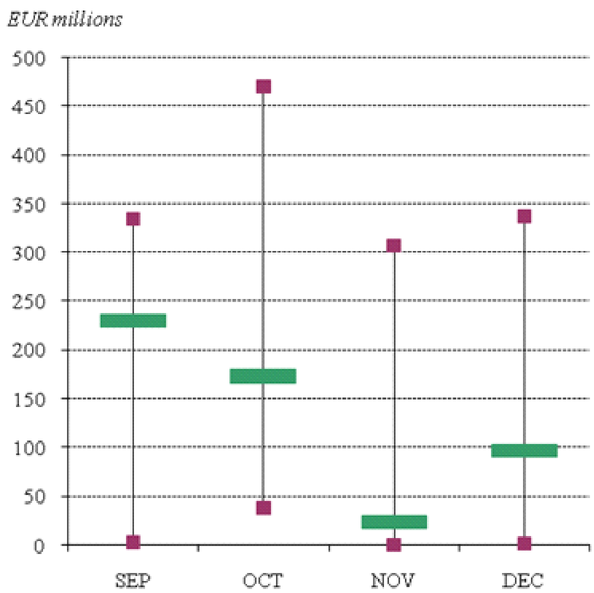
c) The central bank is not exposed to liquidity risk; it can issue currency.

The scenario aims to quantify the impact of halting liquidity injections by the principal participants on queues and unsettled payments. The payments submitted by the State Treasury and central bank were excluded from the sample before running the simulations.

The results for the largest participant indicate both an increased impact as a result of a severe liquidity shock and lower payment-system resilience in September 2008, which is not detected by the medium intensity liquidity shock (Scenario 1).

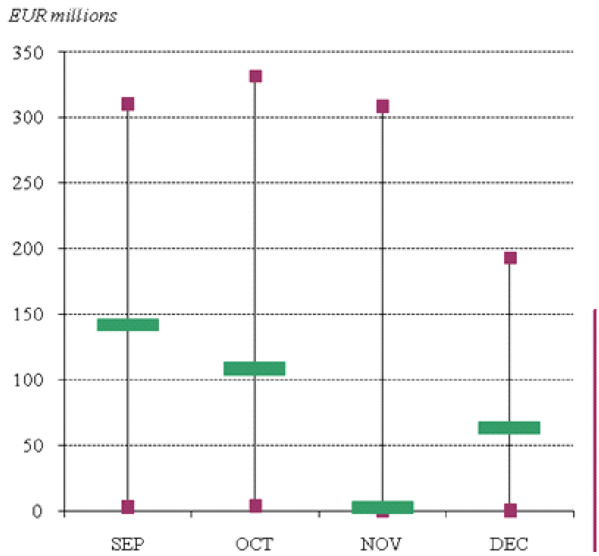
The larger queues in September 2008 (Chart 12.15) may capture some tension in the payment system, which will accelerate and become visible in October 2008. Lehman Brothers bankruptcy in September 2008 sent shock waves through the global financial system and the highly integrated domestic financial system, and the European system soon faced liquidity imbalances. The system was unable to completely absorb the liquidity shock, so at the end of the day there were still unsettled payments. A similar pattern was observed for queues, with the peak in value occurring in September 2008 (Chart 12.16).

Chart 12.15 **Maximum daily queue, September – December 2008**



Source: National Bank of Romania

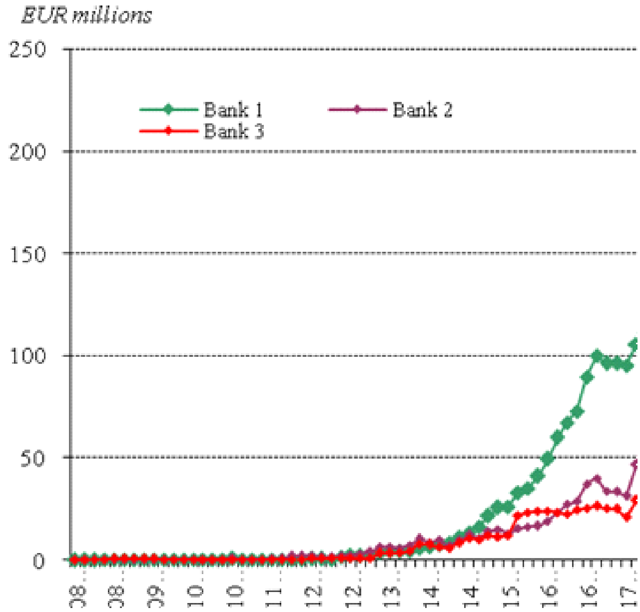
Chart 12.16 **Daily unsettled payments, September – December 2008**



Note: Unsettled payments are end-of-day queued payments
Source: National Bank of Romania

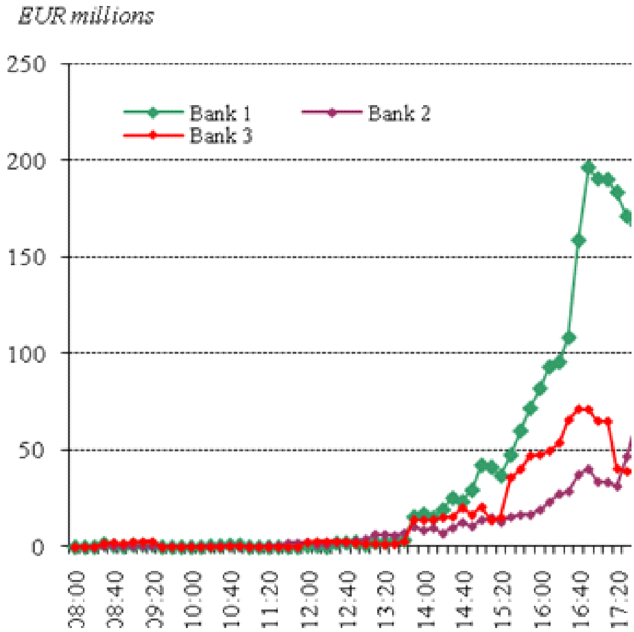
The results for all three largest participants can be observed in charts 12.17 and 12.18. The shock propagation pattern is similar to that of the first scenario. In the first half of the day, the impact is almost nonexistent, but it later increases greatly in intensity. The payment system is not affected by an operational incident if it takes place in the first half of the day and the trading platform is quickly restored; but tensions can arise if the operational incident occurs in the second half of the day.

Chart 12.17 **Shock transmission through the payment system during the day when operational incidents disrupt payment submissions by one of the largest three participants (average values every 10 minutes, September – December 2008)**



Source: National Bank of Romania

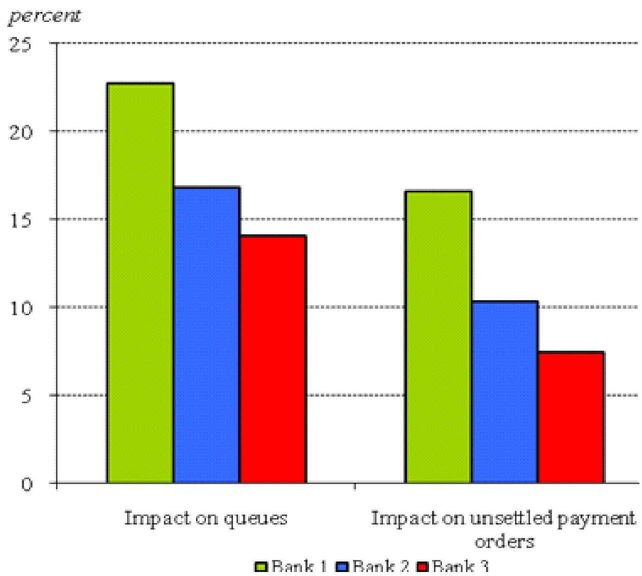
Chart 12.18 **Shock transmission through the payment system during the day when operational incidents disrupt payment submissions by one of the largest three participants (average values every 10 minutes, 13 – 31 October 2008)**



Source: National Bank of Romania

The system cannot fully absorb the liquidity shock when an operational incident affects directly one of the largest three participants, but the magnitude of impact depends on the participant's importance to the payment system. Like the first scenario, the impact increased in 13 – 31 October 2008 due to tensions that already existed in the system, and the size of impact is directly proportional to the participant's importance to the payment system.

Chart 12.19 **The relative impact of transaction not submitted by the first 3 participants on queues and unsettled payment orders, September – December 2008**



Source: National Bank of Romania

The relative size of the impact is also directly proportional to the participant's importance to the payment system. A severe tightening of liquidity injections into the system leads to large values for queues and unsettled payments, but as the shock became less severe, the relative impact decreases significantly (Chart 12.19). Nier, Yang, Yorulmazer and Alentorn (2008) analyzed concentration risk in the banking system and concluded that the greater the concentration, the more exposed is the system to contagion risk when a major participant fails. Their conclusion does not depend on shock magnitude.

12.6 Conclusions

The global financial crisis caused imbalances in the RTGS payment system ReGIS, but not of a great intensity. Liquidity decreased in 13 – 31 October 2008, after Lehman Brothers' bankruptcy sent a shock wave through the global financial system, including the emerging markets, but the situation was stabilized by central bank injections.

Since then, the central bank has become a net creditor of the banking system in terms of market operations.

Based on data and established practices (behavior patterns) of ReGIS, an operational incident, even a severe one, occurring in the first half of the day is quickly neutralized and does not affect liquidity in the payment system because only small value payments are submitted in that part of the day. But tensions can arise in the system if the operational incident occurs in the second half of the day. Queues would increase and some payment orders could remain unsettled by the end of the day, depending on the shock intensity. Even if the shock is completely absorbed, participants are exposed to higher costs of borrowing in the money market, and interest rates may rise.

Improvement in participant abilities to manage liquidity is limited by uncertainty as to the behavior of other participants. The system tends to move to equilibrium, where participants inject into the payment system exactly the amount of liquidity necessary to meet customer needs and avoid liquidity-imbalance signals to other participants. The incentive to accumulate liquidity reserves and to delay payments until funds are received from other participants derives from information asymmetry and cost cutting, which may require a set of options for shifts in behavior.

The central bank, as lender of last resort, provides liquidity to the banking system when needed and can even extend the list of assets accepted as collateral if liquidity is rapidly freezing up. So, the probability that submitted payments remain unsettled at the end of the day is almost zero. This notwithstanding, the current global financial crisis shows that recycling and efficient allocation of liquidity across the system matter more than the volume of liquidity. Liquidity injections provided by central bank allow payment settlements, but imbalances can arise in the payment system, such as rising interest rates and a tightening of the money market, with negative impacts on the real economy.

The study examines the propagation through the payment system of a liquidity shock in the form of an operational incident, rather than an economic or financial crisis. Analyses of liquidity indicators and the system's liquidity allocation mechanism shed light on the system's ability to absorb liquidity shocks, which is directly related to the state of the payment system when an operational incident occurs.

References

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

Glaser, M – Haene, P (2008) **Liquidity effects of a participant-level operational disruption in SIC.** Swiss National Bank.

Koponen, R – Soramaki, K (1998) **Intraday liquidity needs in a modern interbank payment – a simulation approach.** Bank of Finland Studies, No. 14.

Leinonen, H – Soramaki, K (2003) **Simulating interbank payment and securities settlement mechanisms with the BoF-PSS2 simulator.** Bank of Finland, Discussion Paper, No. 23.

Lubloy, A – Tanai, E (2007) **Operational disruption and the Hungarian real time gross settlement system (VIBER).** National Bank of Hungary, Occasional Paper, No. 75.

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

Chapter 13

Does SIC need a heart pacemaker?

Robert Oleschak − Thomas Nellen**

* We are grateful to Beatrice Kraus for valuable research support. We would also like to thank Philipp Haene, Juerg Maegerle, David Maurer, Felix Schafroth, Silvio Schumacher, Dominik Studer and Andy Sturm for helpful comments. The views expressed in this paper are those of the authors and do not necessary represent those of the Swiss National Bank.

13 Does SIC need a heart pacemaker?

Abstract

Real-time gross settlement (RTGS) systems effect final settlement of payments continuously on an order by order basis. This generates a trade-off between liquidity and settlement delay. RTGS systems have thus been enriched with more advanced settlement algorithms aimed at improving the flow of payments and reducing congestion. The paper analyses whether the four most common algorithms can reduce liquidity needs and settlement delay in the Swiss Interbank Clearing (SIC) system. Simulations run with the BoF-PSS2 simulator developed by the Bank of Finland show that expected reductions in delay and liquidity needs are modest and should be evaluated against implementation costs.

13.1 Introduction

Technological innovation, globalization and central bank policies have affected the design of payment systems significantly.[1] In the last three decades, real-time gross settlement (RTGS) systems have emerged and increasingly replaced deferred net settlement (DNS) systems.[2] This development can be understood as a response to the growing awareness of settlement risks in DNS systems.[3] In contrast to DNS systems that accumulate incoming and outgoing payments and settle the net amount at a later, predetermined time, RTGS systems effect

---

[1] For a review on the global trends in large-value payments see Bech, Preisig and Soramäki (2008).

[2] See Bech and Hobijn (2007).

[3] See Bank for International Settlements (BIS) (1997, 2005). Because a DNS system settles the net amount of payments at a later, predetermined time, a lag between the transmission of information about the payment and final settlement occurs. A bank that receives information on a payment that is to be received later, may credit funds to its customer before final settlement occurs. This bank is exposed to settlement risk, as final settlement might not take place as expected. In particular, the bank faces the risk that the system needs to unwind payments if a participant defaults. The unwinding of the defaulting participant's payments affects the end of day balances of receiving participants. This may leave other participants illiquid or insolvent and may thus trigger a domino effect.

final settlement of interbank payments continuously and individually throughout the day.

While final intraday settlement reduces settlement risk, RTGS systems are more liquidity-intensive than DNS systems. In particular, participating banks[4] in RTGS systems face a trade-off between the cost of liquidity and settlement delay. To reduce settlement delay, central banks have typically introduced intraday liquidity facilities, mostly on a free but collateralised basis.[5] Some central banks apply additional measures such as through-put rules or transaction pricing that increases during the day to incentivise early submission and settlement of payments.[6] To further reduce liquidity needs, central banks have aimed to improve RTGS systems' trade-off between liquidity and settlement delay by introducing more advanced settlement algorithms. For instance, the basic first-in first-out (FIFO) algorithm has been enriched by more sophisticated features ranging from simple payment splitting rules and priorities to more advanced features such as bilateral and multilateral offsetting.[7]

The Swiss Interbank Clearing (SIC) system went into service in 1987 and has been operating ever since on the basis of central queuing. When SIC went into operation, payments were settled according to a strict FIFO rule. After minimum reserve requirements were changed in 1988, average reserve balances held by banks dropped drastically (from around CHF 8 billion to CHF 3 billion). As a reaction, the Swiss National Bank (SNB) introduced a two-part tariff and banks together with the SNB agreed to voluntarily split payments that exceed CHF 100 Mio.[8] Since 1994, priorities can be attached to payments. This allows banks to better steer the order of settlement, in particular for time-critical payments. In 2001, SIC introduced a gridlock resolution mechanism that bilaterally offsets payments in case no payments can be settled for a certain period of time.

At the time of writing, discussions on reforming SIC are taking place. One of the issues raised is whether or not the settlement algorithm of SIC should be upgraded. We analyse this question by simulating alternative algorithms based on 'Priorities and FIFO',

[4] Payment system participants may consist of banks as well as other financial intermediaries. For convenience the term banks will be used to refer to all payment system participants.
[5] See World Bank (2008) for a survey of RTGS systems worldwide.
[6] Through-put rules stipulate for participants of a payment system the proportion of daily payments that must be made by a certain cut-off time.
[7] RTGS systems with bilateral or multilateral offsetting algorithms are sometimes referred to as hybrid payment systems since they combine features of RTGS and DNS systems.
[8] Nevertheless, banks frequently settle payments exceeding CHF 100 million.

'Bilateral Offsetting', 'Multilateral Netting' and 'Mandatory Splitting'. For the simulation we rely on real SIC data and on real levels of liquidity. This allows us to effectively compare resulting settlement delay and liquidity usage with the settlement performance of SIC. In doing so, we implicitly assume that banks' submission behaviour and liquidity provision remain constant if a new algorithm is implemented.

Even though SIC operates on the basis of a fairly simple algorithm, we find that more advanced algorithms do not substantially improve the given trade-off between liquidity and settlement delay. We conclude that this is related to the relatively high level of liquidity that allows a simple algorithm to perform well. This complements findings of similar studies for other payment systems.

The observed reductions in delay are very small and are judged to be economically insignificant. First, the reduction in settlement risk due to faster settlement is low. And second, because the cost of intraday liquidity provision is low, the potential reduction of liquidity holdings results in small cost savings. Thus, development and implementation costs should be carefully weighed against these limited benefits when considering the introduction of an advanced algorithm.

Section 13.2 of the paper sets out the theoretical framework, reviews the literature on payment system simulations and presents settlement algorithms used in other countries' payment systems. Section 13.3 describes the settlement algorithm of SIC and funding of payments. Section 13.4 presents the data and the methodology applied. Based on the simulation results presented in section 13.5, a cost-benefit analysis is conducted in section 13.6. The last section closes with concluding remarks.

13.2 Theoretical framework for simulations

13.2.1 Trade-off between liquidity and settlement delay

Banks face a trade-off between liquidity and settlement delay in RTGS systems. A precondition for settlement in RTGS systems is sufficient funding. In case of insufficient funding, submitted payments cannot be settled immediately and will be delayed. Depending on the system design, either payments are rejected and have to be resubmitted or payments are placed in a queue where they are pending until sufficient funding is provided. However, liquidity is costly as reserves held overnight on central bank accounts may yield a lower overnight interest rate. Intraday liquidity is costly too. The central bank determines its price either as an interest rate charge for an uncollateralised overdraft (as applied for example by the Federal Reserve System) or as the opportunity cost of collateral that has to be posted for an interest free intraday credit (as applied for example by the Bank of England (BoE), the Eurosystem and the SNB).[9]

Even though banks can reuse liquidity from incoming payments as a free source of funding for their own payments, timing mismatches between outgoing and incoming payments can cause settlement delay. To save liquidity costs, banks are tempted to free ride on other banks' liquidity, ie banks await incoming payments to fund settlement of their own payments. However, if all banks follow this strategy, negative externalities are created. In particular, excessive delay increases settlement risks as a consequence, for instance, of an operational incident. Furthermore, banks may have to draw excessive levels of liquidity versus the end of the day to be able to settle payments. As pointed out by Angelini (1998), delayed information on incoming payments increases uncertainty and makes liquidity managers hold greater levels of precautionary reserves than they would hold with more precise information on their end of day positions. In addition, delayed settlement may involve pecuniary (such as late settlement fees) or non-pecuniary delay costs (such as the deterioration of a bank's reputation as a reliable trading partner).

Banks deciding between the optimal level of liquidity and settlement delay are further bound by the payment system's transformation curve which is determined by the settlement algorithm. The technical transformation curve is represented in Figure 13.1 by the convex curve AA'.[10] Point A represents a DNS system that settles multilaterally netted amounts at the end of day. As such it is defined as the minimum liquidity necessary to settle all payments at the end of the day with maximum possible delay. In contrast, an RTGS systems can reduce the overall settlement delay but requires additional

[9] Usually, central banks reduce the cost of liquidity by allowing banks to use overnight balances held to fulfil minimum reserve requirements for settlement purposes. The central bank may further seek to reduce the opportunity cost of collateral by accepting a wide range of collateral for intraday credits. Allowing banks to use liquid assets they are required to hold as part of their overnight and longer term liquid asset buffer as eligible collateral for intraday credits – known as double duty – can reduce the opportunity costs of collateral to zero. See Ball, Denbee, Manning and Wetherilt (2011) and Nellen (2012).

[10] The convexity of the curve is based on the assumption of diminishing returns. The higher the level of liquidity, the less reduction of settlement delay results from an additional unit of liquidity.

(intraday) liquidity. Point A' represents the necessary liquidity level to achieve immediate settlement of all payments. RTGS systems usually operate somewhere between the two extremes of the technical transformation curve. Understanding this as a cost minimisation problem, banks try to equilibrate marginal cost of liquidity and delay (as represented by the dashed slope of liquidity cost over delay cost) with the technical rate of substitution. Suppose banks initially end up at point B. If liquidity costs soar (drop), we would expect banks to reduce (increase) their liquidity holdings leading to more (less) delay, moving away from point B up (down) the technical transformation curve.

However, the payment system can also end up above the technical transformation curve – ie a less efficient state. This is for example the case when the synchronisation of payments is not optimal or breaks down due to external factors. For instance, payment systems are vulnerable to disruptions of the payment flow that may also increase uncertainty over the end of day balance position. Such disruptions could be triggered by operational incidences as discussed in McAndrews and Potter (2002). They analyse the effects of the September 11 terrorist attack in 2001 for Fedwire, the US RTGS payment system. Due to widespread damages to property and communication systems the coordination of payments in Fedwire broke down leading to serious settlement delay that could not be managed with the available liquidity. Such disruptions to the synchronisation of the payment flow can move the payment system above the current technical transformation curve. The massive injection of reserve balances during this period was meant to ease liquidity pressure that could have resulted in systemic risk. Essentially, the Federal Reserve System tried to smooth settlement (reduce delay) during the transition from the disruption period to normal operations. With other policies central banks try to affect banks' behaviour to move the payment system's performance close to the technical transformation curve. For instance, in order to reduce the effects of externalities, central banks apply two-part tariffs or through-put rules.[11]

By enriching the settlement algorithm with more advanced features such as bilateral or multilateral payments offsetting may improve the trade-off between liquidity and delay by shifting the

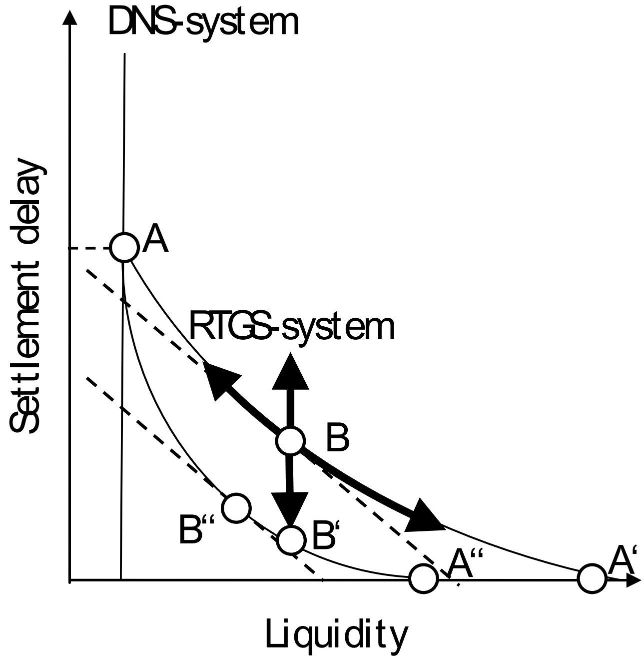
[11] While the SNB for example fosters early release and settlement in SIC by means of a two-part tariff, the BoE induces early release and settlement in CHAPS, the UK RTGS system, by means of a through-put rule. See Ota (2011) for a theoretical discussion on effectiveness of through-put guidelines and two-part tariffs.

transformation curve from AA' to AA''. Banks will choose an equilibrium on (or close) to the transformation curve that is not known to us. Martin and McAndrews (2008) show that settlement algorithms affect banks' behaviour. As a consequence, it is difficult to predict what the new equilibrium and its welfare effects may look like if the settlement algorithm is changed. For instance, simply assuming that the delay and liquidity trade-off remains the same would move us from point B to B''. However, if the trade-off changes, we may end up anywhere on the transformation curve AA'' (or above it).

A simulation study with the Bank of Finland Payment and Settlement System Simulator (BoF-PSS2) on the basis of real data is by definition restricted to measure a move from the original (SIC) to a new technical transformation curve (simulated algorithms) without accounting for potential changes of bank's behaviour. Therefore, we are limited to measure the resulting delay of alternative algorithms under the assumption that the behaviour of banks does not change with regard to chosen liquidity level and release behaviour. To illustrate this, we measure the effect of moving from point B to B', keeping the level of liquidity available and release behaviour the same.

The picture is complicated when available liquidity is split into liquidity that is actually used for settlement (used liquidity) and liquidity that banks may hold for precautionary motives (idle liquidity). As will be shown by the simulations, the level of used liquidity varies for different algorithms for the same amount of available liquidity in a system.

Figure 13.1 **Technical transformation curve between liquidity and settlement delay**



Source: Leinonen and Soramäki (2005), adapted

13.2.2 Simulation results in other countries

Koponen and Soramäki (2005) were the first to conduct simulations with the Bank of Finland-Payment System Simulator (BoF-PSS). Using generated and actual Finnish payment data they find that settlement delay in an RTGS system can be reduced by splitting and netting payments. Splitting of payments worth over 16 million euro has the potential to substantially reduce settlement delay. At very low levels of liquidity splitting may not prevent gridlocks and has limited effects on settlement delay. In contrast, the application of netting is best at reducing settlement delay if the system operates with low levels of liquidity.

These results were later confirmed by simulation studies such as Leinonen and Soramäki (2005) who simulate the effects of splitting and of bilateral and multilateral netting by using real payment data of the Finish RTGS system. They find that settlement delay and the risk of gridlocks – blocked payments in a situation where all participants in aggregate have enough liquidity to settle the end of day multilateral net amount – can be reduced substantially at low levels of liquidity. Going beyond earlier studies, they model banks as economic agents that minimise their private cost of liquidity and delay. Splitting and netting of queued payments is found to reduce the costs of settlement up to 10% for splitting and up to 5% for netting. For both algorithms the cost reduction was most pronounced at low levels of liquidity.

More complex algorithms are, for instance, simulated by Renault and Pecceu (2007). They test the performance of bilateral and multilateral netting algorithms that do not follow the FIFO rule, including the so called GREEDY algorithm proposed by Güntzer et al. (1998). The GREEDY algorithm sorts payments according to value and tries to offset similarly sized payments bilaterally. By using generated as well as real payment data from the Paris Net Settlement System, they show that such algorithms are more efficient than FIFO in terms of unsettled payments for varying levels of liquidity as well as in case of an operational default of a participant. However, Renault and Pecceu (2007) acknowledge that the choice of an algorithm involves other consideration too (for example, an algorithm should be legally sound and match the needs of the users). They conclude that it is difficult to draw definitive conclusion regarding the use of such non-FIFO algorithms in RTGS systems.

Schwinghammer (2002) compares different liquidity-saving mechanisms applied by payment systems in selected countries and discusses the possibility to apply these to SIC. While admitting that there is room for improvement in SIC, Schwinghammer (2002) is sceptical that the benefits of implementing new algorithms in SIC would compensate for the costs. However, Schwinghammer (2002) relies only on the experience of other countries without providing empirical evidence in relation to SIC. Glaser and Haene (2008) simulate the impact on available liquidity if a large SIC participant suffers operational problems and is not able to submit payments but continues on receiving payments for a certain period of time. Because the bank suffering operational problems accumulates liquidity on its account, as a consequence, other participants are hindered to settle due to a lack of liquidity.

13.2.3 Settlement algorithms in different countries

According to a survey by the World Bank (2008), about 80% of the large-value payment systems worldwide are RTGS systems. Most of these systems have a central queuing facility where payment orders are pending until conditions for processing are met. In recent years, increasing numbers of countries have introduced offsetting mechanisms, with multilateral offsetting becoming ever more widely-used. Overall, offsetting algorithms are gaining ground as a means to reduce settlement delay and to save on liquidity in RTGS systems.

Table 13.1 gives an overview of settlement algorithms used in large-value payment systems of the Eurosystem, Japan, Switzerland, the UK and the US. The table shows that a wide range of settlement algorithms are in place ranging from US Fedwire, which has no central queue, to the Eurosystem's large-value payment system TARGET2, which employs arguably the most complex optimisation routines. CHAPS, the British large-value payment system, settles payments according to attached priorities and employs an offsetting algorithm in case a gridlock occurs. This closely resembles the SIC settlement algorithm. Both being so-called hybrid payment systems, the Japanese system, BOJ-NET, and TARGET2 feature continuous bilateral and multilateral offsetting mechanisms.

Because TARGET2 replaced in 2007 and 2008 many national RTGS systems, the effects of its advanced algorithm cannot be assessed. However, BOJ-Net was upgraded with a bilateral and a multilateral offsetting mechanism in October 2008. The accompanying liquidity savings are found to be around 15% (Bank of Japan, 2009).

Table 13.1 **Settlement algorithms in different countries**

System and Country	Basic settlement algorithm	Additional optimisation routines
SIC – Switzerland	Banks can assign priorities to payments. Payments will be ranked according to priority and the first-in first-out (FIFO) principle. Payments are settled in packets, starting with the payments with highest priority.	If no payments can be settled for a certain period (gridlock), a 'circles processing' mechanism is triggered automatically that bilaterally offsets payments. In this case, priority and FIFO are bypassed.
Target2 – Eurosystem	Banks can assign priorities to payments. Highly urgent and urgent payments are settled according to FIFO. Other payments are not settled if highly urgent payments are queued (even if entered first), except where an offsetting transaction of non-urgent payments leads to a liquidity increase for a bank with a highly urgent payment.	Each time a payment is submitted to the system, an offsetting process attempts to bilaterally settle with a payment in the receiving banks queue. Additionally, there are three optimisation routines applied to queued payments. First, an 'all-or-nothing' algorithm tries to settle all payments in the queues simultaneously. If this is not possible, a 'partial' algorithm removes one payment after the other from queues until the remaining payments can be settled simultaneously. Third, a 'multiple' algorithm tries to settle bilateral payments between each pair of banks simultaneously.
CHAPS – United Kingdom	Banks can assign priorities to payments. Within the same priority class, payments with the lowest value are settled first. FIFO is only applied to payments with identical values.	Multilateral offsetting is used if a gridlock occurs.
BOJ-NET – Japan	Banks can manually reorder their queued payments. Bilateral and multilateral offsetting mechanisms can change this order.	A bilateral offsetting mechanism runs continuously while multilateral offsetting is conducted at given time intervals.
Fedwire – United States	As Fedwire does not support central queuing, payments that do not fulfil funding requirements are rejected. However, collateralised or priced overdrafts are granted to ensure smooth settlement.	–

Sources: Bank for International Settlements (2005, Annex 2), Bank of Japan (2009), European Central Bank (2007).

13.3 Settlement and sources of liquidity in SIC

SIC settles large-value payments together with a substantial volume of retail payments in central bank money. Table 13.2 shows that retail payments make up the bulk of payments while large-value payments generate most of the value settled. In 2010, about 380 Swiss and foreign financial institutions participated in SIC and the system handled a daily average of 1.5 million payments with a value of CHF 203 billion. On peak days, SIC processes more than 5 million payments summing up to a settlement value of CHF 425 billion.

Table 13.2 **Share and average size of payments in SIC, 2010**

	Share of transaction volume (In %)	Share of settlement* (In %)	Average payment size (CHF millions)
Large-value payments	4.2	94.1	2.95
Bank to bank	0.9	65.4	9.98
Tri-party repos	0.0	17.8	84.63
SECOM	3.3	10.7	0.43
Eurex	0.0	0.1	2.91
Others	0.0	0.1	11.25
Retail payments	95.8	5.9	0.01
Direct debit	8.3	0.1	0.00
Credit transfer	87.3	5.7	0.01
Batch settlement (POS, cash withdrawals)	0.2	0.1	0.05

* Settlement is defined as the total amount of Swiss francs settled in SIC over one day.
Souce: SNB

13.3.1 Settlement sequence

The exact settlement sequence of payments is determined by the submission behaviour of banks and the settlement algorithm. Payment instructions submitted by a bank are first entered into the waiting queue. If a payment instruction is chosen as settlement candidate and if there is sufficient cover, the payment is settled immediately. If cover is insufficient, the payment remains in the queue until sufficient funding is available. Banks can manage the settlement sequence of their queued payments by assigning priorities to payments.

For payments in the queue, the settlement algorithm determines settlement candidates according to priority classes and the first-in-first-out (FIFO) principle. The process is best explained by differentiating between participant and system level:

- Participant level: As a first step, the settlement algorithm determines the next-highest priority payment to be settled for each bank's queue. If a bank has several payment orders with identical priority in the queue, the payment instruction submitted first will be first in line for settlement.
- System level: If several banks have queued payments, SIC starts to work off the queue of the bank with the longest queue time payment, irrespective of the payment's priority.

For reasons of efficiency, SIC tries to settle several consecutive payments in the same queue. However, the interval of release time and the number of payments that can be settled in one packet are restricted. After settlement took place, the algorithm searches for the next settlement candidate.

If no settlement can be initiated for a certain time interval – a system-wide gridlock – SIC automatically activates a bilateral offsetting mechanism. The mechanism searches for off-setting payments from banks that have sufficient funding for settling the net amount of the two payments. Payments are offset simultaneously on a bilateral basis and, returning to its normal routine, the algorithm searches for the next settlement candidate.

In 1988 a two-part tariff was introduced to encourage early submission and settlement of payments. The remitter of a payment pays a dual-component fee of which one component depends on the time of submission of the payment order and the other one on the settlement time. Both fees increase in the course of the settlement day. The receiver of the payment incurs a flat fee. This tariff scheme creates incentives for banks to submit payment orders early and to provide sufficient liquidity to speed up settlement.

13.3.2 Sources of liquidity

From the viewpoint of a SIC participant, two main sources of liquidity can be distinguished.

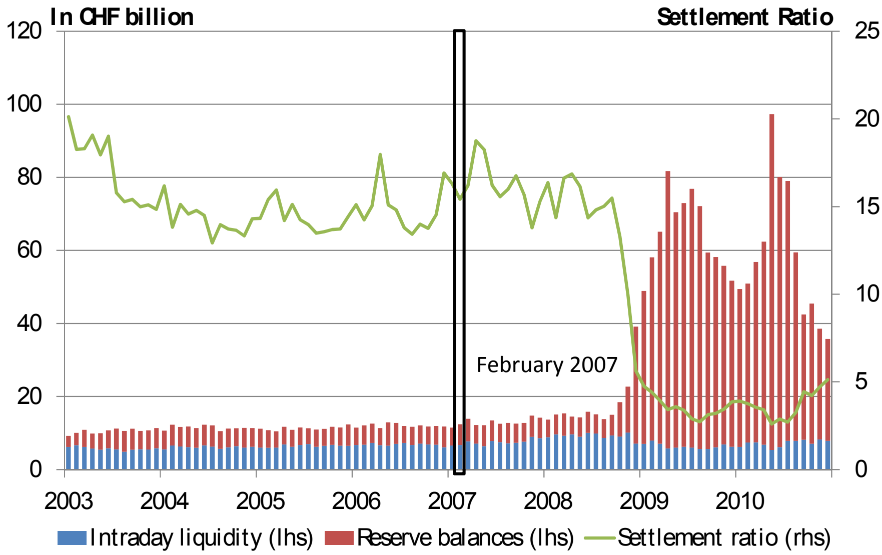
- The first source of liquidity are incoming payments from other system participants. Incoming payments can be used immediately by the receiving participant for the settlement of its own payments.

- The second source of liquidity is the SNB. Each transaction between the SNB and a system participant results in a change of the liquidity available to the participant and to the system as a whole. The SNB distinguishes between open market operations and standing facilities. Open market operations influence the level of overnight reserve balances that banks require to fulfil minimum reserve requirements and that can be used for settlement in SIC. Standing facilities are of importance to settlement purposes only and include the intraday credit facility and the liquidity-shortage financing facility. The intraday-facility provides SIC participants with unlimited interest rate free but collateralised intraday credits during the settlement day. The liquidity-shortage financing facility enables banks at a penalty interest rate to bridge overnight short-term liquidity bottlenecks up to a pre-collateralised limit.[12]

Figure 13.2 depicts monthly averages of total intraday credits drawn over the day, overnight reserve balances and the settlement ratio (SR), which corresponds to the ratio of settlement value to available liquidity (maximum value of intraday credits plus end of day reserve balances). It shows that the level of intraday liquidity, reserve balances and the resulting settlement ratio were pretty stable until the second part of 2008 when unconventional monetary policy measures increased reserve balances to an unprecedented level. The SIC transaction data from February 2007 which is used for our simulations is highlighted in the figure.

[12] See http://www.snb.ch/en/iabout/monpol/id/monpol_instr for further information on the SNB's monetary policy instruments.

Figure 13.2 **Monthly average of intraday liquidity, reserve balances and settlement ratio**



Source: SNB

13.4 Data and simulation methodology

This section describes the data used, the simulation algorithms applied, and the methodology for measuring liquidity and settlement delay.

13.4.1 Data sample

We conduct simulation on the basis of data from February 2007 because transaction volumes and values represent average SIC activity. Furthermore, February 2007 is a pre-crisis month that represents an average level of liquidity in normal times (see Figure 13.2). The sample covers 15 business days with an average daily number of 1.2 million transactions and an average daily settlement value of 167 billion Swiss francs.

The BoF-PSS2 Simulator works on the basis of a 24 hour settlement day. Because SIC starts the settlement day for Mondays on Friday 5 p.m. and ends it – with two interruptions on Saturday and

Sunday – on Monday 4.15 p.m., we exclude Mondays from the sample. We further exclude 28 February because the volume was exceptionally high.

We further extract CLS related transactions because these payments are settled on specifically dedicated subaccounts that do not influence settlement on the main accounts. This is also related to the funding of these subaccounts that is done exclusively via intraday credits. Because these intraday credits are often transferred to the main accounts after the CLS settlement day is finished, we account for liquidity movements that take place between main accounts and CLS subaccounts as these affect settlement.[13]

13.4.2 Alternative settlement algorithms

Given the variety of settlement algorithms in use, we focus on the most generic forms of algorithms selecting the features most often applied (see Table 13.3):

- The first algorithm 'Priority and FIFO' serves as a reference case. It is the BoF-PSS2 algorithm that represents the closest available approximation of the algorithm currently applied in SIC.
- Based on the basic 'Priority and FIFO', the second algorithm applies a continuous 'Bilateral Offsetting' mechanism. This additional mechanism checks – for each new payment order submitted − whether the payee has an approximately offsetting payment waiting in the queue that is directed towards the payor. If positive, the payments are offset by the settlement of the net amount.
- The third algorithm complements the second algorithm with a 'Multilateral Netting' every 60 minutes. It tries to settle all payments in the queue simultaneously. If the multilateral net settlement does not succeed, it reverts to 'Bilateral Offsetting'.
- The fourth algorithm improves the first algorithm by introducing 'Mandatory Splitting' of payments larger than CHF 100 million. While other amounts could be chosen, in the case of SIC it is most interesting to analyse a limit of CHF 100 million. This allows us to

[13] See Appendix B for a detailed description of how these transactions were excluded and how liquidity movements between CLS subaccounts and main accounts are taken into account.

evaluate the effects if the currently applied voluntary splitting becomes mandatory.

Table 13.3 **Simulation algorithms**

Number and label	Basic settlement algorithm	Additional optimisation routine
1. 'Priority and FIFO'	Payments are queued if liquidity is insufficient. Payments are released according to priority and FIFO if liquidity becomes available.	–
2. (1.) + 'Bilateral Offsetting'	Same basic settlement algorithm as 'Priority and FIFO'.	Continuous bilateral offsetting is applied that can bypass strict system level priority FIFO order transactions.
3. (1.)+(2.)+ Full 'Multilateral Netting' every 60 minutes	Same basic settlement algorithm as 'Priority and FIFO'.	In addition to continuous bilateral offsetting, complete multilateral netting takes place every 60 minutes on the basis 'all or nothing'.
4. (1.) + 'Mandatory Splitting' of transactions greater than CHF 100 million	Same basic settlement algorithm as 'Priority and FIFO'.	Transactions larger than CHF 100 million are split.

The basic settlement algorithm of the BoF-PSS2 Simulator 'Priority and FIFO' resembles the SIC settlement algorithm. However, three differences cannot be replicated with the available built in logics of BoF-PSS2 simulator:[14]

– Selection of a bank's queue: If there are several banks with a payment queue in the system with sufficient funds to settle payments, the SIC algorithm starts settling the queue first which contains the payment that has been submitted first, irrespective of its priority. In contrast, 'Priority and FIFO' settles the queue first which contains the payment with the highest priority. 'Priority and

[14] Algorithms of BoF-PSS2 can be adapted and modified by user if needed. In this study, however, the built in algorithms were used as such and considered to resemble the real SIC logic accurately enough despite the listed differences.

FIFO' resorts to FIFO only in case several queues with identical priorities exist.

- Packet building: Once SIC has chosen a queue, it continues to settle all payments within the same priority until one of the following conditions are met: the participant's queue is empty, the maximum volume of payments within a given limit is settled (one packet should not contain more than 150 payments), the maximum time lag between the first and last payment is reached (currently set at one minute) or cover becomes insufficient. In contrast, 'Priority and FIFO' tries to settle all payments in the chosen queue until cover becomes insufficient.
- Gridlock resolution mechanism: In case no queued payments can be settled for a certain period of time, SIC activates the gridlock resolution mechanism. 'Priority and FIFO' does not have such a gridlock resolution algorithm in place.

To compare the simulation results of the four alternative algorithms described above with the current SIC algorithm, we calculate the delay and liquidity indicators for SIC too. Applied liquidity and delay indicators are described in the following section.

13.4.3 Measuring liquidity and delay

Available liquidity in the SIC system is equal to the liquidity provided by the SNB. We define available liquidity (LA) to be equal to the sum of overnight balances at the end of the day plus the sum of all intraday credits drawn during the day. Banks can draw and pay back intraday liquidity at any time after 7.30 am on. Thus, LA could vary during the day. Also, the repurchase leg (purchase leg) of open market operations take place after 8 am (9 am). However, the suggested definition for LA is reasonable for the following two reasons. First, banks almost don't vary their holdings of intraday credits during the day.[15] Second, available overnight balances at end of day reflect available overnight liquidity during the hours of the greatest settlement activity. Thus, we define LA as follows:

$$LA = B(t_M) + \sum_i d^i(t_0, t_M) \tag{13.1}$$

[15] See Nellen (2012).

where $i = 1,2,\ldots,N$ (number of participants) and $m = 0,1,2,\ldots,M$ (number of time intervals); $B(t_M)$ represents the balance of all participants at the end of day (t_M); and $d^i(t_0,t_M)$ represents the sum of intraday credits drawn by bank i between beginning (t_0) and end of day (tM).[16]

Available liquidity can be divided into liquidity that is actually used to effect settlement (used liquidity, LU) and liquidity that remains idle on the accounts of banks (idle liquidity, LI). LI can be derived via simulation. It is defined as the sum of reserves that lie idle on the accounts of banks, ie the sum of the minimum account balances in the course of the settlement day.[17] Thus, LU is the residuum:

$$LU = LA - LI \tag{13.2}$$

Independent of the reasons why individual banks hold more liquidity than they actually use to effect settlement of their payment obligations (such as precautionary reserve holdings, or different levels of sophistication of liquidity management), we treat the observed level of idle liquidity in SIC as a minimum level that is not further reduced even if more advanced algorithms would allow to do so. As a consequence, if advanced settlement algorithms result in an increase of idle liquidity, the difference between idle liquidity for SIC and idle liquidity for a more advance algorithm could be eliminated and represents potential liquidity savings.

For each measure of liquidity a ratio can be calculated, dividing the respective measure of liquidity by the settlement value. The available liquidity ratio (ALR) is of particular interest as it is often used as a reference for the liquidity efficiency of a payment system:

$$ALR = \frac{LA}{(\text{Settlement Value})} = \frac{1}{SR} \tag{13.3}$$

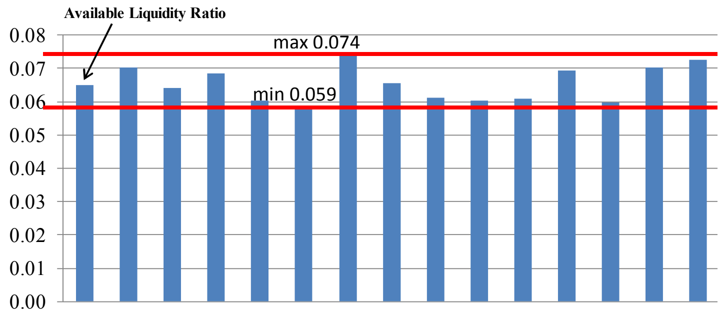
As we use data from before the financial crisis starting 2007, we implicitly assume that – after the effects of the financial crisis have subdued – the level of available liquidity and the value settled will return to similar levels. Therefore, we presume that the available

[16] See Appendix A for a more detailed description and derivation of liquidity measures.
[17] Some algorithms end the day with pending payments (see Table 13.4). In order to make the levels of idle liquidity comparable, the value of unsettled payments (if any) is always deducted from idle liquidity.

liquidity ratio – as a crucial driver of settlement speed – returns to the same stable relationship in the future (see Figure 13.2).

Even though we do not artificially change the level of available liquidity, natural day to day fluctuations of available liquidity and settlement value allow us to investigate the effects of changing levels of liquidity to a certain degree. During the observation period the ALR varied between 0.059 and 0.074, corresponding to a SR between less than 14 and more than 17 (see Figure 13.3).

Figure 13.3 **Available liquidity* (CHF billion) and available liquidity ratio in SIC**



Day in Feb 2007	01.	02.	06.	07.	08.	09.	13.	14.	15.	16.	20.	21.	22.	23.	27.
Balance End-of-Day	4.7	5.6	5.1	5.8	5.4	5.7	5.8	5.5	5.4	5.6	6.0	6.6	6.1	6.3	5.9
Intraday Repos	5.7	5.8	4.5	4.7	5.4	4.1	5.6	5.1	5.6	4.8	5.1	4.9	5.2	5.7	5.6
Available Liquidity	10.4	11.4	9.6	10.5	10.8	9.7	11.4	10.6	10.9	10.4	11.1	11.5	11.3	12.1	11.5
Idle Liquidity	1.2	1.0	1.2	1.2	1.0	0.9	1.1	1.2	1.2	1.1	1.1	1.2	0.9	1.0	1.2
Available Liquidity Ratio	0.07	0.07	0.06	0.07	0.06	0.06	0.07	0.07	0.06	0.06	0.06	0.07	0.06	0.07	0.07

*Excluding transactions that have been used for making CLS payments.

We measure delay with two standard settlement delay indicators[18] supported by the BoF-PSS2 Simulator:

$$\text{Settlement Delay Weighted (SDW)} = \frac{\sum_k^K q_k \times a_k}{\sum_k^K p_k \times a_k} \tag{13.4}$$

$$\text{Settlement Delay Unweighted (SDU)} = \frac{\sum_k^K q_k}{K} \tag{13.5}$$

[18] For a detailed description consult the BoF-PSS User Manual (User Manual: Databases and Files, Appendix 13.1).

where k represents the number of payments of all banks k=0,1,2,…,K, q_k represents queuing time for payment k, a_k represents value of payment k, and p_k represents maximum settlement delay, ie the time difference between submission and end of day of payment k.

While the indicator 'settlement delay weighted' (SDW) weighs queuing time of each payment with its value, the indicator 'settlement delay unweighted' (SDU) simply represents the average queuing time of all payments irrespective of their value. Furthermore, SDW assigns queued payments a higher weight the later in the day they are queued. This comes as a result of the divisor weighing a payment according to its potential queuing time which is defined as end of day time minus submission time. To illustrate the difference between the two indicators consider a payment system that runs for 23 hours. A single payment of one unit of money is to be settled. If the payment is submitted to the system at the beginning of the day and is queued for one hour, the weighted settlement delay indicator for t equals 1/23 = 0.043. In contrast, the weighted settlement delay indicator equals 1/1 = 1 if the payment is queued for one hour but submitted at t=22. For comparison, the SDU of both payments' is equal to 60 minutes. Therefore, SDU is neutral with regard to the time of queuing whereas SDW assigns more weight to higher value payments and delays taking place later in the day.[19]

13.5 Simulation results

The results of the four simulated algorithms ('Priorities and FIFO', 'Bilateral Offsetting', 'Multilateral Netting' and 'Mandatory Splitting') for 15 days in February 2007 are summarised in Table 13.4 below. Based on these results and by comparing them to the settlement delay and used liquidity observed for the SIC algorithm we identify the following key findings (detailed statistics are reported in Appendix 13.C).

On average additional liquidity shows a limited effect on settlement delay. However, additional liquidity reduces delay for high-

[19] The weighted settlement delay indicator may be motivated by respective preferences. For instance, early release of payments is considered beneficial to the system as a whole as it may speed up settlement and reduces operational bottlenecks at the end of the day. Also, queues later in the day may bear higher delay costs. For instance, should an operational outage occur later in the day, it is associated with greater settlement risk. One may interpret a SDW of 0.11 as follows: on average each Swiss franc was queued 11% of the time between release and end-of-day.

value payments later in the day more effectively. The indicators for settlement delay unweighted (SDU) and settlement delay weighted (SDW) exhibit negative correlations with the available liquidity ratio (ALR) and used liquidity ratio (ULR). This indicates that additional settlement value adjusted liquidity translates into lower delay. While the correlations are low for SDU, correlations for SDW are higher. This indicates that additional liquidity allows to reduce delay of high-value payments that are submitted later in the day.

The SIC algorithm seems to be on average superior in speeding up settlement of low-value payments (retail payments) compared to any other simulated algorithm with same level of available liquidity. While we observe a slightly higher level of SDW for the SIC algorithm compared to all other simulated algorithms, the SIC algorithm is superior in terms of SDU. This suggests that the SIC algorithm is better at settling the high share of low-value payments than any other algorithm simulated in this study. This, however, comes at a cost of higher volatility: for both delay measures the SIC algorithm exhibits a higher standard deviation than the simulated algorithms (see Tables 13.6 and 13.7 in Appendix 13.C).

The simulations show that compared to 'Priorities and FIFO' more advanced settlement algorithms do not allow to reduce settlement delay substantially. This holds true for any given level of available liquidity observed. 'Bilateral Offsetting' reduces average SDU by 9% and average SDW by 12% compared to 'Priorities and FIFO'. There is no added-value of 'Multilateral Netting' compared to 'Bilateral Offsetting' since it neither reduces delay nor used liquidity. There are two possible explanations for this observation. First, the 'Multilateral Netting' algorithm works on the basis 'all or nothing'. If the algorithm cannot settle all queued payments at once, it reverts to the 'Bilateral Offsetting' modus. It seems that there were no or only a few situations where multilateral netting of all queued payments succeeded. Secondly, the two largest participants in SIC account on average for about 50% of all payments in terms of value (see Glaser and Haene, 2008) and most of these payments occur bilaterally between those two participants. This particular participation structure limits the usefulness of the 'Multilateral Netting' compared to the 'Bilateral Offsetting' algorithm further. Even though 'Mandatory Splitting' (with a limit of CHF 100 million) reduces settlement delay compared to ‘Priorities and FIFO’, it does so to a negligible extent (1.3% reduction for SDU and 0.3% reduction for SDW).[20]

For ‘Bilateral Offsetting’ and ‘Multilateral Netting’ no unsettled payments remain pending at the end of the day, unlike for ‘Priorities and FIFO’ and ‘Mandatory Splitting’ where payments are left unsettled. ‘Priorities and FIFO’ exhibits on average 35 payments with a value of almost CHF 300 million unsettled. ‘Mandatory Splitting’ ends the day on average with 14 unsettled payments with a value of CHF 250 million. In comparison to an average settlement value of CHF 167 billion and average volume of 1.2 million payments, pending payments at the end of the day remain negligible. Therefore, indicators of delay are not materially affected by the value and volume of pending payments.

Available liquidity (LA) is derived from the liquidity that banks have received in February 2007 from the SNB during the 15 days we used for the simulation. To ensure comparison between the SIC algorithm and the four simulated algorithms, the level of LA has been kept the same. However, differences in idle liquidity (LI) and, consequently, the level of used liquidity (LU) arise:

- Compared to ‘Priorities and FIFO’, the SIC algorithm exhibits a higher level of LI of about CHF 100 million on average. Thus, showing an almost identical SDW but much lower SDU, the SIC algorithm seems to be more efficient than ‘Priorities and FIFO’.[21]
- Compared to ‘Priorities and FIFO’, all simulated settlement algorithms are more liquidity efficient since they all have higher levels of idle liquidity and show lower delay indicators. Banks use CHF 270 million less liquidity in case of continuous ‘Bilateral Offsetting’ and ‘Multilateral Netting’, while ‘Mandatory Splitting’ increases idle liquidity to a much lesser extent.

Because the more advanced algorithms are not based on the SIC algorithm but are built on the ‘Priorities and FIFO’ algorithm, it is also interesting to compare the more advanced algorithms with ‘Priorities and FIFO’. Compared to ‘Priorities and FIFO’, we find ‘Bilateral Offsetting’ to reduce settlement delay and liquidity needs to a limited degree. Furthermore, we take the unchanged performance of ‘Multilateral Netting’ as an indication that – at the given levels of

[20] In February 2007 on average around 150 transactions with a share of 15% in total settlement value were above CHF 100 million.
[21] In order to make the levels of idle liquidity comparable, the value of unsettled payments (if any) is always deducted from idle liquidity.

liquidity – the added-value of more sophisticated algorithms is limited. This impression is corroborated by the evidence found for 'Mandatory Splitting', which reduces settlement delay and liquidity usage very moderately.

Table 13.4 **Simulation results and comparison to SIC, February 2007**

	SIC	Simulation results			
	Original	1. Priorities and FIFO	2. Bilateral Offsetting	3. Multilat. Netting	4. Mandatory Splitting
Average SDW#	0.157	0.153	0.135	0.135	0.151
Corr SDW/ALR	-0.28	-0.54	-0.44	-0.44	-0.50
Corr SDW/ULR	-0.30	-0.54	-0.33	-0.33	-0.39
Average SDU##	15.18	19.40	17.70	17.70	19.35
Corr SDU/ALR	-0.21	-0.28	-0.22	-0.22	-0.28
Corr SDU/ULR	-0.21	-0.22	-0.19	-0.19	-0.27
Average Available Liquidity+	10,884	10,884	10,884	10,884	10,884
Average Used Liquidity+	9,778	9,882	9,608	9,608	9,799
Average Idle Liquidity+	1,106	1,002	1,275	1,275	1,085
Liquidity safed+	104	-	274	274	83
Number of unsettled payments	nap§	35	0	0	14
Value of unsettled payments++	nap§	296	0	0	248

An interpretation of the settlement delay weighted (SDW) with value 0.16 is that each Swiss franc was on average in the queue 16% of the time between its submission and end-of-day. ALR stands for available liquidity ratio and ULR for used liquidity ratio.
Settlement delay unweighted (SDU) measures the average queuing time of a payment (in minutes).
+ In CHF million.
++ Benchmark is the settlement algorithm "Priorities and FIFO". Higher levels of idle liquidity indicate, that less liquidity was used for actual settlement (in CHF million).
§ Number and value of unsettled payments is by definition zero, since SIC system deletes all payments that remain in the queue by the end of the day. Details of deleted payments are not recorded in the databank.

13.6 Cost-benefit analysis

A comprehensive cost-benefit analysis is beyond the scope of this study. However, we identify potential sources of benefits and costs associated with introducing a new algorithm.

The potential sources of benefits include the reduction of delay and liquidity needs. The simulation has shown that 'Bilateral Offsetting' reduces in comparison to 'Priorities and FIFO' both settlement delay weighted (SDW) by 9% (or by 0.018 on range between 0 and 1) as well as settlement delay unweighted (SDU) by 12% (or by 1.7 minutes per payment). The relatively small reduction of delay in terms of both indicators suggests that related benefits are economically insignificant. Furthermore, for a large-value payment system that is also used as a settlement engine for a high volume of retail payments, one should carefully consider the objectives a new algorithm should serve. The results indicate that SIC is able to accommodate the need to timely settle large-value payments as well as to cope with a large volume of retail payments.

The second source of benefits are potential cost savings resulting from the reduction of liquidity holdings. Assuming that banks hold of the same level of idle liquidity as seen in February 2007 to cope with unexpected payments, any increases to the level of idle liquidity associated with the introduction of a new algorithm can then be set equal to reductions of liquidity needs. Using 'Priorities and FIFO' as the benchmark algorithm and assuming that adding 'Bilateral Offsetting' to the SIC algorithm reduces liquidity needs by the same margin, we find that 'Bilateral Offsetting' reduces liquidity needs of banks by around CHF 280 million. As reserve balances are mainly held to fulfil minimum reserve requirements, participants would reduce this liquidity by means of lowering their demand for intraday credits. Since intraday credits are free but collateralised, the potential cost savings are calculated by multiplying the average liquidity saving with the implicit intraday interest rate. For the period after the introduction of CLS in 2002 and before the financial crisis 2007, Kränzlin and Nellen (2010) estimate the implicit intraday interest rate to be around 2.7 basis points. Therefore, yearly cost savings due to a lower provision of intraday credits would be about CHF 75,600 (CHF 280 Mio x 2.7 basis points).

The introduction of a new algorithm involves costs related to its development and implementation. In case of SIC – settling both large-value and retail payments – an off-the-shelf algorithm may not be the appropriate choice. However, a customised solution increases development costs. In addition, a new algorithm may give rise to substantial adaption costs for participants. For instance, adaption costs could be caused by the need to rearrange internal payment processing arrangements. Besides these sunk costs, a new algorithm may have to be carefully design to avoid higher variable costs related to an increasing demand for processing capacity and management attention.

13.7 Conclusions

The paper investigates whether the trade-off between delay and liquidity in SIC can be improved with the introduction of more advanced algorithms aimed at improving the flow of payments and reducing congestions. Using real transaction and liquidity data, we compare the performance of the current SIC algorithm with the performance of the simulator's basic algorithm 'Priorities and FIFO' as well as with three more advanced algorithms, namely 'Bilateral Offsetting', 'Multilateral Netting' and 'Mandatory Splitting'.

We find that compared to 'Priorities and FIFO', 'Bilateral Offsetting' is able to modestly reduce delay and liquidity usage. Interestingly, 'Multilateral Netting' – which is built on top of 'Bilateral Offsetting' – provides no value added. Enriching 'Priorities and FIFO' with 'Mandatory Splitting' reduces settlement delay and liquidity usage to a negligible degree.

Assuming that adding 'Bilateral Offsetting' to the SIC algorithm reduces delay and liquidity usage by the same margin as it does if added to 'Priorities and FIFO', we find potential yearly liquidity costs savings of around CHF 75,000 as a result of reduced needs of intraday liquidity. Measured reductions in settlement delay are considered to be economically insignificant. The costs associated with the introduction of a new algorithm such as investment, adaption and running costs must be weighed against these benefits.

Our findings are in line with other studies. Sophisticated settlement algorithms reduce delay and liquidity usage substantially only if the level of liquidity is low. The level of available liquidity in SIC is sufficient to ensure smooth settlement and does not leave room for sophisticated algorithms to take effect. Submission and settlement timing has improved considerably over the past 20 years. In April 1988, the introduction of two-part tariffs incentivised early submission and settlement and fostered an efficient sequencing of payments that allows to process large volumes of retail payments together with large-value payments. Delay was substantially reduced after the introduction of interest-free intraday credits in 1999. Due to the increase in settlement liquidity by about CHF 8 billion, delay was almost eliminated by 2007. Therefore, the relatively minor effects of alternative settlement algorithms in SIC are not surprising.

Overall, we find that the current SIC algorithm performs comparatively well and that alternative algorithms offer only very modest improvements of the trade-off between liquidity and delay. These benefits are thus very likely to be outweighed by the costs associated with the development, implementation and adaption of a new algorithm.

Appendix A

Measures for available, used and idle liquidity

Let $B^i(t_m)$ represent the balance of bank i at time t_m. The balance is equal to the balance at the beginning of day (BoD), plus the difference between the cumulative value of outgoing and incoming payments until t_m from and to other banks (s), overnight repos and any other flows between participant i and central bank (o) or intraday repos received from or paid back to the central bank (d).

$$\begin{aligned} B^i(t_m) = BoD^i + \textstyle\sum_{j \neq i}[s^{ji}(t_0,t_m) - s^{ij}(t_0,t_m)] + [o^{ci}(t_0,t_m) \\ - o^{ic}(t_0,t_m)] + [d^{ci}(t_0,t_m) - d^{ic}(t_0,t_m)] \end{aligned} \tag{A13.1}$$

where $i = 0,1,2,\ldots,N$ (number of participants) and $m = 1,2,\ldots,M$ (number of time intervals). $s^{ij}(t_0,t_m)$ = settled payments from bank i to bank j between t_0 and t_m. $o^{ci}(t_0,t_m)$ = settled overnight or longer repos and other flows from central bank c to bank i between t_0 and t_m. $d^{ci}(t_0,t_m)$ = settled intraday repos from central bank c to bank i between t_0 and t_m.

The liquidity available (LA) in the payment system at time t_m equals the sum of balances of all system participants. Note that interbank payments cancel out, thus LA is defined as

$$\begin{aligned} LA(t_m) = B(t_m) = \textstyle\sum_i BoD^i + \sum_i [o^{ci}(t_0,t_m) - o^{ic}(t_0,t_m)] \\ + \textstyle\sum_i [d^{ci}(t_0,t_m) - d^{ic}(t_0,t_m)] \end{aligned} \tag{A13.2}$$

The maximum available liquidity is defined as

$$MaxLA = \max_m (B(t_m)) \tag{A13.3}$$

Maximum available liquidity as defined in equation (A13.3) is not easily detectable as banks can draw and pay back intraday liquidity at any time during the day (but latest by the end of the day). Banks have to settle the repurchase leg of maturing overnight repos before they can draw new ones, which takes place at around 9 a.m. Assuming that banks typically pay back their intraday liquidity holdings after conducting their overnight repo transactions, maximum liquidity available can be approximated by the sum of the end-of-day balance and the peak intraday liquidity position

$$\text{Max LA}_{\text{approx}} = B(t_M) + \sum_i d^{ci}(t_0, t_M) \qquad \text{(A13.4)}$$

$$\begin{aligned} B(tM) &= \sum_i BoDi + \sum_i o^{ci}(t_0, t_M) - o^{ic}(t_0, t_M)] \\ &= \text{Balance end of day} \end{aligned} \qquad \text{(A13.5)}$$

Liquidity available in a system can be divided into liquidity used (LU) and liquidity that has been lying idle on the account and has not been used for making payments (LI). Therefore we have

$$LA(t_M) = LU(t_M) + LI(t_M) \qquad \text{(A13.6)}$$

By definition, an individual bank's LI is the stock of funds on its settlement account that could be siphoned off without any effect on the bank's payment performance at any time of the day. Overall, the system can settle each bank's payments in exactly the same manner with or without banks' individual shares in LI on their accounts. Given this definition, LI in a system can be defined as the sum of idle liquidity holdings over all banks

$$LI(t_M) = \sum_i (\min_m (B^i(t_m))) \qquad \text{(A13.7)}$$

Inserting (A13.4) and (A13.7) in (A13.6) we get a measure for maximum liquidity used in the system

$$\text{MaxLU}(t_M) = B(t_M) + \sum_i d^{ci}(t_0, t_M) - \sum_i (\min_m (B^i(t_m))) \qquad \text{(A13.8)}$$

Appendix B

Ancillary systems and treatment of CLS-payments

SIC is linked to the securities settlement system SECOM, to the central counterparty Eurex Clearing and to the foreign exchange settlement system CLS. The payments resulting from SECOM and Eurex Clearing are left unaltered for the purpose of the simulation analysis. A few other ancillary systems settle in participants' main accounts on the basis of direct debit payments. These are left unaltered too. CLS-related transactions are removed as such payments settle in dedicated CLS sub-accounts and do not affect delay or liquidity needs on the main accounts. However, transactions between main accounts and CLS sub-accounts are replaced by SNB related payments in order to replicate liquidity implications for the banks. If cash is transferred from a bank's main account to CLS, a corresponding payment debiting the bank's main and crediting the SNB's account is created. If cash is transferred from CLS to a bank's main account, a corresponding payment crediting the bank's main account is created. All other CLS related payments are removed from the transaction data. Figures A13.1 and A13.2 illustrate the procedure.

Figure A13.1 **Transactions with CLS sub-account**

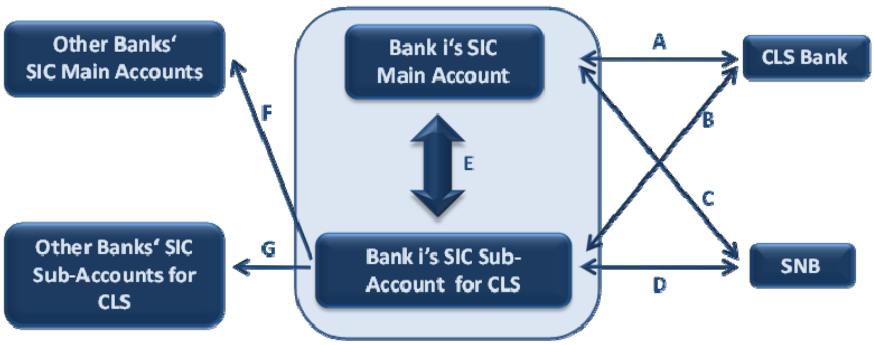
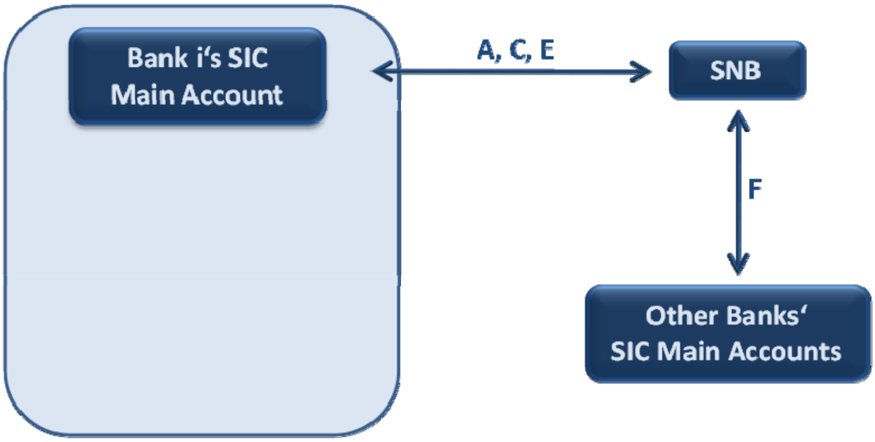


Figure A13.2 **Transactions without CLS sub-account**



Appendix C

Data

Table A13.1 **Value and Volume of Transactions for days used in simulation***

Date	Value of transactions (in Mio CHF)	Volume of transactions
01.02.2007	160,219	2,549,714
02.02.2007	162,082	1,527,960
06.02.2007	149,469	1,150,163
07.02.2007	152,838	1,008,863
08.02.2007	179,214	844,145
09.02.2007	166,378	986,297
13.02.2007	154,112	808,995
14.02.2007	161,266	772,400
15.02.2007	179,153	869,732
16.02.2007	172,167	795,222
20.02.2007	182,239	818,271
21.02.2007	165,385	718,179
22.02.2007	189,285	837,156
23.02.2007	171,289	2,197,914
27.02.2007	158,349	2,006,079
Average	166‘896	1,192,739

* CLS-Transactions and beginning of day transactions are not included.

Table A13.2 **Simulation results – Settlement Delay Weighted**

Day	Available Liquidity Ratio	Settlement Delay Weighted				
		SIC	Priorities and FIFO	Bilateral Offsetting	Multilateral Netting	Mandatory Splitting
01.02.2007	0.066	0.20	0.16	0.14	0.14	0.16
02.02.2007	0.070	0.17	0.15	0.14	0.14	0.15
06.02.2007	0.065	0.15	0.14	0.12	0.12	0.14
07.02.2007	0.069	0.17	0.17	0.15	0.15	0.17
08.02.2007	0.061	0.17	0.17	0.15	0.15	0.17
09.02.2007	0.058	0.14	0.14	0.12	0.12	0.14
13.02.2007	0.074	0.14	0.14	0.12	0.12	0.14
14.02.2007	0.065	0.17	0.17	0.15	0.15	0.17
15.02.2007	0.061	0.18	0.18	0.16	0.16	0.17
16.02.2007	0.060	0.14	0.16	0.13	0.13	0.16
20.02.2007	0.060	0.13	0.15	0.13	0.13	0.14
21.02.2007	0.070	0.13	0.13	0.12	0.12	0.12
22.02.2007	0.060	0.18	0.18	0.16	0.16	0.18
23.02.2007	0.070	0.15	0.15	0.13	0.13	0.15
27.02.2007	0.072	0.11	0.11	0.10	0.10	0.11
Average	0.065	0.155	0.153	0.135	0.135	0.151
St. deviation	–	0.026	0.020	0.017	0.017	0.020

Table A13.3 **Simulation results – Settlement Delay Unweighted, in Minutes**

Day	Available Liquidity Ratio	Settlement Delay Unweighted				
		SIC	Priorities and FIFO	Bilateral Offsetting	Multilateral Netting	Mandatory Splitting
01.02.2007	0.066	30.44	25.50	24.77	24.77	25.45
02.02.2007	0.070	21.47	20.33	18.60	18.60	20.28
06.02.2007	0.065	8.93	13.07	12.15	12.15	13.05
07.02.2007	0.069	10.61	13.68	11.83	11.83	13.12
08.02.2007	0.061	17.15	22.75	21.10	21.10	22.77
09.02.2007	0.058	9.62	14.93	13.78	13.78	14.70
13.02.2007	0.074	7.80	15.93	16.10	16.10	15.72
14.02.2007	0.065	13.69	27.88	21.23	21.23	28.03
15.02.2007	0.061	21.30	26.58	24.55	24.55	26.52
16.02.2007	0.060	13.88	17.38	16.22	16.22	17.38
20.02.2007	0.060	16.77	21.50	18.12	18.12	21.87
21.02.2007	0.070	13.73	14.72	13.55	13.55	14.70
22.02.2007	0.060	17.02	21.42	19.48	19.48	21.43
23.02.2007	0.070	11.77	15.13	14.28	14.28	15.13
27.02.2007	0.072	13.53	20.22	19.70	19.70	20.12
Average	0.065	15.18	19.40	17.70	17.70	19.35
St. deviation		5.89	4.86	4.17	4.17	4.96

Table A13.4 **Simulation results – Number of Unsettled Transactions**

Day	SIC*	Priorities and FIFO	Bilateral Offsetting	Multilateral Netting	Mandatory Splitting
01.02.2007	nav	0	0	0	0
02.02.2007	nav	11	0	0	11
06.02.2007	nav	0	0	0	0
07.02.2007	nav	7	0	0	7
08.02.2007	nav	14	0	0	14
09.02.2007	nav	30	0	0	30
13.02.2007	nav	0	0	0	0
14.02.2007	nav	23	0	0	23
15.02.2007	nav	61	0	0	61
16.02.2007	nav	14	0	0	14
20.02.2007	nav	17	0	0	17
21.02.2007	nav	4	0	0	4
22.02.2007	nav	7	0	0	7
23.02.2007	nav	19	0	0	19
27.02.2007	nav	314	0	0	10
Average	0	35	0	0	14

* Unsettled payments in SIC are removed from the payment statistics and hence not available.

Table A13.5 **Simulation results – Value of Unsettled Transactions, in CHF million**

Day	SIC*	Priorities and FIFO	Bilateral Offsetting	Multilateral Netting	Mandatory Splitting
01.02.2007	nav	0	0	0	0
02.02.2007	nav	393	0	0	393
06.02.2007	nav	0	0	0	0
07.02.2007	nav	1	0	0	149
08.02.2007	nav	266	0	0	266
09.02.2007	nav	403	0	0	403
13.02.2007	nav	0	0	0	0
14.02.2007	nav	393	0	0	393
15.02.2007	nav	311	0	0	311
16.02.2007	nav	569	0	0	569
20.02.2007	nav	348	0	0	348
21.02.2007	nav	45	0	0	45
22.02.2007	nav	83	0	0	83
23.02.2007	nav	386	0	0	386
27.02.2007	nav	1241	0	0	379
Average	0	296.04	0	0	248.42

* Unsettled payments in SIC are removed from the payment statistics and hence not available.

References

Angelini, P (1998) **An analysis of competitive externalities in gross settlement systems.** Journal of Banking & Finance 22, 1–18.

Renault, F – Pecceu, J-B (2007) **From PNS to TARGET 2: to cost of FIFO in RTGS payment systems.** Bank of Finland, Scientific monographs, E:39, 41–85.

Ball, A – Denbee, E – Manning, M – Wetherilt, A (2011) **Intraday Liquidity: Risk and Regulation.** Bank of England, Financial Stability Paper, No. 11.

Bank for International Settlements (1997) **Real-time Gross Settlement Systems.** Bank for International Settlements, Basel. http://www.bis.org/publ/cpss22.pdf?noframes=1 (29 March 2010).

Bank for International Settlements (2005) **New developments in large-value payment systems.** BIS/CPSS, May 2005), http://www.bis.org/publ/cpss67.pdf?noframes=1 (29 March 2010).

Bank of Japan (2009) **BOJ-Net Funds Transfers after the Implementation of Phase 1 of the Next Generation RTGS Project.** Bank of Japan Review, July 2009.

Bartolini, L – Hilton, S – McAndrews, J (2008) **Settlement delays in the money market.** Federal Reserve Bank of New York Staff Report no. 319.

Bech, M – Hobijn, B (2007) **Technology diffusion within central banking: the case of real-time gross settlement.** International Journal of Central Banking, Vol. 3 No. 3, 147–181.

Bech, M – Preisig, C – Soramäki, K (2008) **Global trends in large-value payments.** Federal Reserve Bank of New York Economic Policy Review, September 2008, 59–81.

European Central Bank (2007) **Guideline of the European Central Bank of 26 April 2007 on a Trans-European Automated Real-time Gross settlement Express Transfer system (TARGET2).** http://www.ecb.int/ecb/legal/pdf/02007o0002-20091123-en.pdf (31 March 2010)

Glaser, M – Haene, P (2008) **Liquidity effects of a participant-level operational disruption in the Swiss Interbank Clearing System.**

Güntzer, M M – Jungnickel, D – Leclerc, M (1998) **Efficient algorithms for the clearing of interbank payments.** European Journal of Operational Research, Vol. 106, Issue 1, 1 April 1998, 212–219.

Koponen, R – Soramäki, K (2005) **Intraday Liquidity Needs in a Modern Interbank Payment System – A Simulation Approach.** Bank of Finland Studies E:31, 73–116.

Kränzlin, S – Nellen, T (2010) **Daytime is money.** Journal of Money, Credit and Banking, Vol. 42, No. 8.

Leinonen, H – Soramäki, K (2005) **Optimising Liquidity Usage and Settlement Speed in Payment Systems.** Bank of Finland Studies E:31, 117–150.

Martin, A – McAndrews, J (2008) **Liquidity-Saving Mechanisms.** Journal of Monetary Economics 55, No. 3 (April): 554–567.

McAndrews, J – Potter, S M (2002) **Liquidity Effects of the Events of September 11, 2001.** In: FRBNY Economics Policy Review, November 2002, 59–79.

Mills, D C Jr. – Nesmith, T D (2007) **Risk and Concentration in Payment and Securities Settlement Systems.** Journal of Monetary Economics 55, 542–553.

Nellen, T (2012) **Double duty and two-part tariff.** Mimeo, Swiss National Bank.

Ota, T (2011) **Intraday two-part tariff in payment systems.** Bank of England Working Paper No. 428.

Schwinghammer, C (2002) **Liquiditätssparende Verrechnungs-algorithmen.** Analyse und Schlussfolgerungen. Swiss Interbank Clearing AG. Version 1.1.

World Bank (2008) **Payment Systems Worldwide. A snapshot.** Outcomes of the global payment system survey 2008. World Bank Financial Infrastructure and Research Series.

Chapter 14

The impact of payment system design on tiering incentives

Robert Arculus – Jennifer Hancock* − Greg Moran**

* We are grateful to Mark Manning, Andrea Brischetto, Christopher Kent, other colleagues at the Reserve Bank of Australia, and participants at Bank of Finland's 2009 Payment and Settlement System Simulation Seminar for their helpful comments. We would also like to thank Sammy Yousef, Rod Phillips and the Bank of Finland staff for their technical advice. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Reserve Bank of Australia. The authors are solely responsible for any remaining errors.
Authors: hancockj or morang at domain rba.gov.au
Media Office: rbainfo@rba.gov.au

14 The impact of payment system design on tiering incentives

Abstract

Tiering – where an institution does not participate directly in the central payment system but instead settles its payments through an agent who does – is a significant issue for payment system regulators. Indirect settlement can provide efficiency advantages, particularly in terms of liquidity savings, but it can also increase risk. This paper uses simulation analysis to explore the impact of payment system design on institutions' incentives to tier. Some evidence is found to support the hypothesis that the liquidity-saving mechanisms in Australia's real-time gross settlement (RTGS) system, the Reserve Bank Information and Transfer System (RITS), reduce the liquidity cost of direct participation and thus may have contributed to the low level of tiering in RITS relative to RTGS systems in other countries. The paper also attempts to quantify the increases in concentration and credit risk that would occur if there were to be an increase in the extent of tiering in RITS from current low levels, and the effect of system design on credit risk. Increased tiering is found to result in only small increases to the level of concentration in RITS, though it could lead to more substantial increases in the share of total payments that the larger individual institutions are responsible for processing (that is, both through the central system and across their own books) when acting as an agent for a number of smaller institutions. In terms of credit risk, increased tiering would create substantial two-way exposures between clients and their settlement banks, although there is no clear relationship between system design and the size of these exposures. Finally, the paper considers a number of issues related to evaluating the relative benefits and costs of tiering.

14.1 Introduction

Most central payment systems settle high-value payments on an RTGS basis. This prevents the build-up of large interbank exposures, which would otherwise occur if high-value payments were settled on a deferred net basis. However, RTGS systems require participants to hold substantial liquidity in order to make payments. Where participants access liquidity via repurchase agreements, they incur an opportunity cost because the collateral posted cannot be used for other purposes.

Tiering – where an institution does not participate directly in the central payment system but instead settles its payments indirectly through an agent that does – is a significant issue for payment system regulators. On the one hand, tiering can reduce system liquidity needs because:

- payments between a tiered participant (client) and its settlement bank are settled across the settlement bank's books rather than sent to the central system; and
- combining the payment flows of the client(s) with those of the settlement bank may allow the settlement bank to fund more payments from receipts rather than from liquidity accessed via repurchase agreements.

On the other hand, tiering can increase both credit and concentration risk. Tiering, by definition, increases concentration in the system as more payment activity occurs through a smaller number of direct participants. Credit risk also arises because the settlement bank and its client are each exposed to the failure of the other. These benefits and costs of tiering, and others, are discussed further in Section 14.2.

The degree of tiering varies across payment systems. The CHAPS system in the United Kingdom, for instance, is relatively highly tiered, with only 17 direct participants (not including the Bank of England) making payments on behalf of several hundred other institutions (Bank for International Settlements, 2012). In contrast, the US Fedwire system has a fairly flat payments structure, with several thousand direct participants. Australia's RTGS system, RITS, also has a low level of tiering. While in the early days of RITS this was due to restrictions on tiering, these restrictions were relaxed in 2003 to allow institutions whose RTGS payments are less than 0.25 per cent of the total value of RTGS payments to settle through an agent. Since then, however, very few institutions have opted to settle indirectly. In 2008, around half of RITS's 67 participants were below the 0.25 per cent threshold and therefore eligible to settle indirectly, yet only six chose to do so. Given that the vast majority of eligible participants joined RITS prior to the relaxation of tiering restrictions, it is possible that moves to indirect settlement have been deterred in part by the prior absorption of fixed costs associated with becoming a direct participant, or simple organisational inertia. However, the low level of tiering does raise the question of what drives participants' incentives to tier and whether aspects of system design reduce the incentive to tier in RITS.

This paper uses simulation analysis to explore the impact of payment system design on institutions' incentives to tier. Specifically, it tests the hypothesis that including liquidity-saving mechanisms in the design of an RTGS system reduces the incentives to use tiering to save liquidity. It also attempts to quantify the increases in concentration and credit risk that would occur if there were an increase in tiering in RITS from current low levels, and the effect of system design on credit risk. Finally, it discusses the relevant considerations in weighing-up the estimates of the benefits and costs of tiering. This analysis is intended to shed light on the present level of tiering in RITS, as well as inform policymakers in regard to rules that restrict tiering.

The remainder of the paper is structured as follows. Section 14.2 briefly reviews the literature on the costs and benefits of tiering in payment systems. Section 14.3 provides an overview of RITS and Section 14.4 outlines the simulation methodology. Based on these simulations, Section 14.5 presents estimates of liquidity savings from tiering under different system designs. Section 14.6 presents estimates of the increases in concentration and credit risk that would occur if there was to be an increase in tiering. Section 14.7 discusses how the benefits and costs of tiering might be weighed. Section 14.8 concludes.

14.2 The benefits and costs of tiering

14.2.1 Benefits

Systems that operate on an RTGS basis require participants to hold substantial liquidity in order to cover payments as they arise. In RITS, intraday liquidity is provided through interest-free repurchase agreements ('repos') with the Reserve Bank of Australia (RBA), but participants incur an opportunity cost as collateral posted using this facility is unavailable for alternative uses. As discussed in Jackson and Manning (2007), Adams, Galbiati and Giansante (2010) and Lasaosa and Tudela (2008), tiering can reduce the liquidity needs of an RTGS system because:

– payments between the client and the settlement bank are settled across the settlement bank's books, rather than being sent to the RTGS system (payments internalisation)
– combining payment flows allows more payments to be funded from receipts (liquidity pooling). Unless the client's and settlement bank's peak intraday liquidity requirements occur simultaneously, tiering requires less liquidity than the sum of their individual peak requirements since payments received by one can be used to fund payments by the other.

While saving on liquidity is the potential benefit of tiering in which we are primarily interested in this paper, several other benefits are identified in the literature. Jackson and Manning (2007) and Adams et al (2008), for instance, explore the idea that tiering can benefit a system if some participants have lower costs of direct participation than others. Chapman, Chiu and Molico (2008) and Kahn and Roberds (2008) suggest that tiering encourages inter-agent monitoring of credit worthiness, which may be more efficient than monitoring by the central bank or requiring the posting of collateral.

14.2.2 Potential impact on risk

While there are potential benefits from tiering in payment systems, there can also be costs. In particular, tiering can increase a number of types of risk in a payment system. Perhaps the most significant of these is credit risk. Just as moving to an RTGS system decreases credit risk at the expense of increased liquidity costs (see Kahn and Roberds, 2008), tiering represents the possible reintroduction of credit risk. Note that this credit risk is two-way. Both the settlement bank and its client are exposed to the failure of the other; the former because it may offer its client intraday credit and the latter due to the settlement bank's role as holder of the relevant accounts. Harrison, Lasaosa and Tudela (2005) attempt to quantify the credit exposure of settlement banks in CHAPS, finding that the risk is not substantial under normal operating conditions, but has the potential to rise considerably in extreme circumstances. To manage this change in credit risk, settlements banks may well react by reducing the credit they extend to their client banks in times of stress. This 'liquidity dependence' may have a significant effect on the indirect participant as it no longer has direct access to central bank liquidity.

Tiering can also increase concentration risk. An operational problem at a participant, for instance, may result in that participant becoming a 'liquidity sink'.[1] The more liquidity is concentrated into fewer participants, the more severe the impact of such a problem is likely to be. On the other hand, as a tiered network depends less on the central infrastructure, it may allow some payments to still go ahead in the event of a central system failure (although not in central bank money). The net effect is ambiguous, but certainly tiering has the potential to significantly alter the effects of system disruptions and participant failures.

While the focus in this paper is on credit and concentration risk, other risks that can arise from tiering include:

- the legal risk that the finality of payments settled in commercial bank money is not protected in the same way as the finality of payments settled in central bank money[2]
- the business risk that the exit of a settlement bank from the market may cause a greater disruption to the payments system than would result were tiering not present
- the competitive risk involved in a settlement bank also being a competitor with its clients in the market for retail payment services (see Chande, Lai and O'Connor, 2006).

14.3 Australia's RTGS system

RITS has operated as an RTGS system since 1998.[3] The central queue in RITS operates on a 'bypass first-in first-out (FIFO) basis'.[4] If the transaction being tested for settlement cannot be settled individually, the bilateral offset algorithm searches for up to 10 offsetting transactions (additively, in FIFO order), which it attempts to settle

[1] This is a situation where a participant is able to receive but not send payments, and thereby drains liquidity from the system.
[2] For instance, under the 'zero hour' rule, a court may date the bankruptcy of an institution from the midnight before the bankruptcy order is made. In Australia, the Payments Systems and Netting Act 1998 allows the RBA to protect payments that occur in RITS from the application of this rule, but payments settled across the books of a settlement bank do not have the same protection.
[3] For more information on RTGS in Australia see Gallagher P, Gauntlett, J and Sunner, D (2010).
[4] Payments are tested for settlement in FIFO order, but if a payment fails the settlement tests the system moves on to test the next payment in the queue for settlement, and so on, looping back to the first payment when it reaches the end of the queue.

simultaneously.[5] RITS incorporates other queue management features, which allow participants to manage their payments and reserve liquidity for 'priority' payments. To assist in this process, RITS participants have access to real-time information, including their settled and queued payments and receipts. The liquidity reservation feature in RITS allows participants to set a 'sub-limit', with balances below this limit reserved for the settlement of payments flagged as having 'priority' status. Payments flagged as having 'active' status are tested for settlement against balances in excess of the sub-limit, while payments flagged as 'deferred' are not tested for settlement until the sending participant changes the status of the payment to either active or priority. Participants can amend the status of payments at any time prior to settlement.

Connection to RITS is via either the internet or infrastructure shared with the Australian debt securities depository and settlement system, Austraclear. The RBA does not charge for internet connections to RITS. Thus, non-liquidity costs of direct participation in this case are those associated with equipment, office space, staff training and salaries, and internet service provision. Membership of Austraclear involves initial and annual fees. If an institution is already a member of Austraclear and chooses to connect to RITS through the Austraclear infrastructure, then the incremental non-liquidity costs of direct participation are just those pertaining to staff. In general, these costs are likely to vary considerably across institutions and are difficult to estimate accurately.

Initially, direct access to RITS was only available to banks, with all banks required to settle their RTGS payments using their own settlement account.[6] In 1999, following the recommendations of the Wallis Inquiry into Australia's financial system, access was broadened to allow third-party payment providers and non bank Authorised Deposit-taking Institutions (ADIs) to participate directly in RITS.[7] The Wallis Inquiry also resulted in the creation of the Australian Prudential Regulation Authority (APRA), which regulates all ADIs – banks, building societies, credit unions and special third-party providers of payments services. While all ADIs can now become

[5] In July 2009, the RBA added a Targeted Bilateral Offset algorithm, which allows participants to select specific payments for bilateral offset.

[6] Special Service Provider accounts were set up for the building society and credit union industry associations, to allow building societies and credit unions to settle indirectly through these associations.

[7] See Reserve Bank of Australia (1999) for more information.

direct participants in RITS, only banks are *required* to hold a settlement account at the RBA.

Notwithstanding the broad scope of participation, payments through RITS are highly concentrated, with the major domestic banks accounting for almost 60 per cent of the value of all payments made. Indeed, payments just between the four major domestic banks account for around a third of all payments. Also, the direction of payment flows tends to be skewed. For example, most RITS participants make more than half of their payments, by value, to just a few other participants.

Since 2003, settlement account-holders whose RTGS transactions individually comprise less than 0.25 per cent of the total value of RTGS transactions have been permitted to settle indirectly via an agent.[8] Prior to this, account-holders were prohibited from tiering. Despite the relaxation in policy, however, data for 2008 show that only six of the 34 participants eligible to settle indirectly chose to do so.

14.4 Methodology

Our methodology is adapted from Lasaosa and Tudela (2008), who study the benefits and costs of tiering in CHAPS using the Bank of Finland's payment system simulator, BoF-PSS2. The simulator models the operation of RTGS systems (or other large-value payment and settlement systems) described by a set of parameters and data. Key inputs include transaction data and credit limits (used to model the liquidity available to participants in the system). The simulator's output includes the settlement profile of payments and measures of the liquidity used by participants in the system.

Lasaosa and Tudela create tiering scenarios for simulation by amending raw transaction data from CHAPS. For example, to model Bank A settling indirectly through Bank B, they create an amended transaction dataset in which payments originally to or from Bank A become payments to or from Bank B. Payments originally between Bank A and Bank B are deleted from the data, as these are now settled across Bank B's books rather than submitted to the system. These now 'internalised' payments are an immediate source of liquidity savings.

[8] See Australian Prudential Regulation Authority and Reserve Bank of Australia (2003) for more information.

We create tiering scenarios by amending transaction data from RITS in the same way. The sample period is the month of January 2008, covering 21 business days over which 623 860 individual transactions took place with a total value of around $4.04 trillion. Excluding a number of participants for which indirect settlement would be unrealistic (such as the four largest participants, CLS Bank and the RBA), there are 49 participants altogether that are considered candidates for tiering in this experiment. Note that only the smallest 25 of these 49 candidates were under the 0.25 per cent threshold in 2008 and therefore eligible to tier.[9] Notwithstanding this, we model both the cumulative effect of all participants below a given size settling indirectly (the 'cumulative scenarios'), and each of the 49 candidates individually electing to tier (the 'individual scenarios'), resulting in 98 unique sets of transaction data representing 98 unique tiering scenarios.[10]

As Lasaosa and Tudela are primarily interested in the effect of a decrease in tiering in a highly tiered system, they used the results of their simulations to forecast this effect. Given the high level of participation in RITS, such forecasting was unnecessary in the context of this paper.

It should be noted that the analysis here is necessarily limited in that it ignores the potential for payments behaviour of participants to change in response to different tiering arrangements. Because the transaction and credit limit inputs to the simulator specify, inter alia, payment submission times, payment statuses (eg priority or active) and maximum liquidity accessible, none of these can be optimised by participants in response to different levels of tiering.

14.4.1 Tiering order

Although there are a number of ways to select client institutions and their respective settlement banks (see Lasaosa and Tudela for examples), we allocate institutions based on the value of payments sent and received. In the cumulative scenarios, the 49 candidates are tiered from smallest to largest in order of their share of all payments. Our reasoning is that larger institutions will generally have a lower

[9] The 28 direct participants eligible for tiering over the whole of 2008 include one participant not considered for tiering in our simulations and two participants that joined RITS after January 2008.

[10] In the cumulative scenarios in which the fifth-largest institution is tiered, the 49 smallest institutions are all settling indirectly via the four largest participants.

opportunity cost of collateral as their banking operations naturally result in their holding more eligible securities on their balance sheet, which in turn gives them a competitive advantage in the market for providing payment services. This approach is also consistent with the current formulation of RBA policy, whereby only participants whose share of RTGS payments comprise less than 0.25 per cent of the total value of RTGS transactions are eligible to tier.

The settlement bank for each individual tiering candidate is chosen as the institution with which the candidate conducts the largest share of its payments. This approach is likely to maximise the value of payments that are internalised, although this is not a mathematical certainty.[11]

In practice, such allocation decisions would be interdependent. That is, each institution's choice of settlement bank could change depending on the choices of other institutions and the subsequent sizes of different tiered networks (Adams et al provide an interesting model of participant tiering choice). However, preliminary work suggested that attempting to account for this would have minimal effect; for instance, when each client institution was assigned to its largest payments partner with the choices of all smaller institutions taken as given, the choice of settlement bank differed only on four occasions.

14.4.2 System design

To test the hypothesis that the liquidity-saving features of RITS decrease participants' incentives to tier, we simulate tiering in four RTGS system designs (Table 14.1). Details of how the bilateral offset and sub-limit features of RITS have been incorporated in the simulations are contained in Appendix A.

Table 14.1 **RTGS system designs**

	Central queue	Bilateral offset	Sub-limits
Pure RTGS	–	–	–
RTGS with central queue only	x	–	–
RTGS with bilateral offset	x	x	–
RITS replica	x	x	x

[11] A further scenario, based on the order of the share of total volumes, was not materially different to the one based on values, and so it was not pursued further.

Regardless of our tiering order methodology, we expect liquidity use to fall as we increase the number of liquidity saving mechanisms in the system. That is, we expect liquidity use to be the greatest under the pure RTGS system, followed by central-queue-only, then bilateral offset. The RITS replica is expected to require the lowest level of liquidity.

14.4.3 Liquidity

The liquidity available to participants is modelled in the simulations using limits on credit extended by the system operator to each direct participant. Each participant begins each simulated day with an account balance of zero and, as payments settle, is able to accrue a negative account balance up to that participant's credit limit. Credit limits are set exogenously and may vary throughout the day. In general, the credit limit profile for each participant on each simulated day is modelled on the actual liquidity that was available to that participant at each point in time on the corresponding day of our sample period. This actual liquidity is measured as the sum of the participant's opening balance and the value of intraday repos it had outstanding at each point in time during the day.[12]

An exception is made for our simulation of the pure RTGS system. To prevent payments that do not settle immediately from being rejected and remaining unsettled at the end of the day, all participants are assumed to have access to unlimited liquidity. In addition, to ensure that all payments settle in our simulations, we give all participants unlimited access to liquidity under all system designs at the end of each day.[13]

In the tiering scenarios, we reason that the settlement bank does not have access to collateral on its clients' balance sheets and it will not necessarily commit more of its own collateral to access additional liquidity. Alternatively, we could have assumed that the settlement bank increased the liquidity it accesses (for example, by the value of the liquidity accessed by its clients when they were direct

[12] In our experiments, the RBA, CLS Bank and the settlement accounts of the equity and futures clearing and settlement systems are provided with unlimited credit in all system designs.

[13] In the absence of this we find that the simulations result in a small proportion (less than 1 per cent) of payments remaining unsettled at the end of most days. This failure to settle all payments occurs because settlement times differ across the different RTGS systems, while available liquidity is set exogenously.

participants). Indeed, preliminary simulations were run where the credit limits of the settlement bank and its clients were summed, but this resulted in quite substantial and unrealistic increases in liquidity usage under tiering. Therefore, our preference has been to remain with fixed, non-additive access to liquidity.

We measure system liquidity usage as the sum of individual participants' peak intraday liquidity requirements. For an individual participant, this peak intraday liquidity requirement is equal to the absolute value of the participant's minimum account balance. While this liquidity may only have been used for a very brief period during the day, this measure is consistent with our belief that the main driver of the cost of liquidity is the maximum value of collateral used, rather than the length of time during the day that the securities are used.

14.5 The impact of tiering on liquidity usage

14.5.1 Estimates of liquidity savings

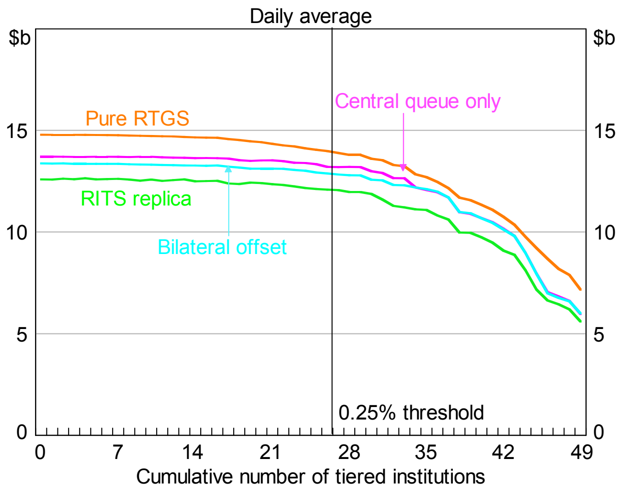
We present simulation results for the changes in liquidity use due to increased tiering in the four system designs. Our hypothesis is that the liquidity benefits from tiering decrease as more liquidity saving features are added to the RTGS system. A decomposition of liquidity savings into the two sources identified in the literature, namely liquidity pooling and payments internalisation, is contained in Appendix B.

14.5.1.1 Cumulative tiering

We first look at the case where individual institutions are tiered cumulatively, from smallest to largest, according to their share of the total value of payments. Figure 14.1 shows liquidity usage over the range from no tiering to tiering all candidate institutions. For all scenarios in this case, the pure RTGS system is the most liquidity intensive and the RITS replica the least intensive. Of the two other system designs, the bilateral offset system clearly uses less liquidity for the first 33 tiering scenarios. For subsequent scenarios, however, the presence of bilateral offset has almost no effect. This may be due to the increasing concentration of the system; Ercevik and Jackson (2009) report the intuitive finding that liquidity recycling increases with system concentration, thus the need for bilateral offset decreases.

The share of the total value of payments settled by bilateral offset falls from 28 per cent when there is no tiering to 13 per cent when all candidate institutions are tiered.

Figure 14.1 **System liquidity usage**



Source: RBA

In line with our hypothesis, liquidity-saving mechanisms typically reduce the liquidity benefits from tiering in this case. Average daily liquidity usage falls by $7.0 billion when all candidates are tiered in the RITS replica system, compared with larger decreases of: $7.4 billion in the system with bilateral offset; $7.6 billion in the pure RTGS system; and $7.7 billion in the system with a central-queue-only. However, there is some variance in the results, with this ranking not holding at all increments of the cumulative tiering.[14]

14.5.1.2 Tiering individual participants

We now look at the case where individual participants are tiered in isolation. Again, the pure RTGS system is the most liquidity intensive and the RITS replica the least intensive for all scenarios. Average daily system liquidity usage falls by $137 million on average when a

[14] For the systems with credit limits, we find that the marginal change in liquidity usage from tiering an additional institution is often not significantly different from zero at the 10 per cent level for approximately the smallest 30 institutions.

single institution is tiered in the RITS replica system, compared with larger decreases of: $143 million in the system with a central-queue-only; $151 million in the pure RTGS system; and $155 million in the system with bilateral offset. While liquidity savings are lowest in the RITS replica system as expected under our hypothesis, the fact that savings are highest in the bilateral offset system is not consistent with our hypothesis. Again, there is variance in the results, with this ranking not necessarily holding for each individual institution tiered.

14.5.1.3 Network effects

It is possible that the low level of tiering observed currently in RITS might be a result of the benefits of tiering being dependent on the size of the tiered network. For instance, the proportion of payments that can be internalised for a given tiering candidate will tend to increase the larger is the tiered network being joined. If these network effects are large, then multiple equilibria including both high and low degrees of tiering would be conceivable (with the latter a result of very few institutions considering it worthwhile to tier as long as very few other institutions are tiered already).

By comparing the liquidity savings in the cumulative and individual tiering scenarios, we find some evidence for this possibility. When the smallest 30 institutions are tiered in the RITS replica system, liquidity usage declines by around 5.7 per cent. When each of the smallest thirty institutions are tiered individually, the summed marginal changes imply a decrease in liquidity usage of around 3.9 per cent (that is, the effect of tiering institutions simultaneously accounts for around one-third of total liquidity savings). However, this estimate of the size of the network effects varies; if we considered tiering 35 institutions, for instance, then network effects would appear to account for only 10 per cent of liquidity savings. In addition, on average an institution in the smallest 30 sends and receives just 8 per cent of the total value of its payments to and from other institutions in the smallest thirty, suggesting that the scope for network effects in this group is limited.

14.6 The impact of tiering on risk

The benefits of tiering can come at a cost of increased concentration and credit risk. This section estimates the changes in concentration and credit risk in RITS due to increased tiering. The effect of system design on credit risk is also examined.

14.6.1 Concentration risk

Indirect participants in a payments network send payment instructions to their settlement bank, which then acts on their behalf. Consequently, in choosing to tier the client becomes operationally dependent on its settlement bank. One might argue that larger institutions are better equipped to minimise the probability of an operational problem. However, by concentrating payment flows, tiering amplifies the consequences of an operational incident at the settlement bank – in particular, the size of the potential liquidity sink increases.

A general measure of this type of operational risk is the level of concentration in the system: the increase in settlement banks' share of payments as the level of tiering increases. Note that our measure of concentration is the share of payments sent, rather than sent and received, as generally even when a participant suffers an operational incident they can still receive payments. While a more accurate way to model the impact of tiering on the consequences of an operational incident is to simulate operational incidents in a tiered network, this is beyond the scope of this paper.

We find that our cumulative tiering scenarios result in only a modest increase in the concentration of payments being sent to RITS by the four largest participants. In the absence of tiering, the four largest participants account for around 57 per cent of all payments sent to RITS by value. If all of our 49 tiering candidates were to settle indirectly the combined share of the four largest participants would rise by around 10 percentage points. While this is an indication of the liquidity dependence of indirect participants on the remaining settlement banks, since it is unlikely that an operational incident would occur at all four of the largest participants simultaneously, it is more noteworthy that the largest single share only rises 4 percentage points.

An alternative measure of concentration is the value of payments that the four largest participants are collectively responsible for; that is, the value of payments sent by them to the central system plus the value of payments settled across their books. By this measure the rise in concentration is more substantial, at just over 24 percentage points. In addition, the largest single share rises by around 12.5 percentage points. Thus the extent to which concentration risk is an issue depends on the relative likelihoods one attaches to different types of operational outages; that is, whether outages are more likely to simply affect the ability of an institution to access the central system, or whether they are more likely to disrupt the processing of payments entirely. We do not pursue this issue further here.

14.6.2 Credit risk

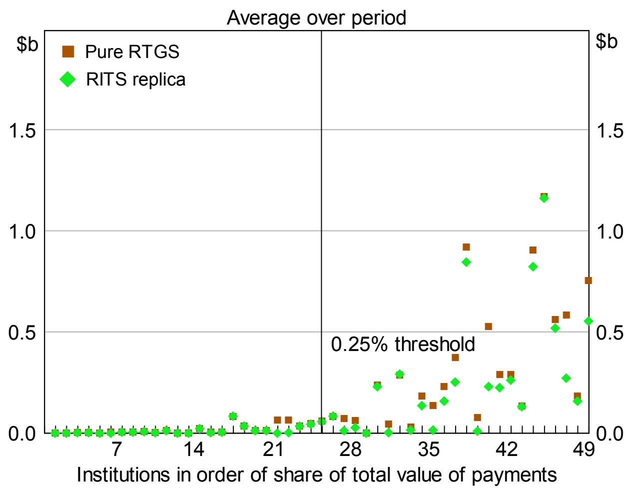
Tiering creates a two-way exposure between a client and its settlement bank because payments are settled across the settlement bank's books, rather than in central bank money (which is not subject to credit risk). Furthermore, these payments – unlike those in RITS – may be subject to the zero hour rule, which means that in the event of a bankruptcy, their finality can be challenged. In this section we present measures of this two-way exposure for the two system designs at either end of the liquidity usage spectrum: the pure RTGS system and the RITS replica system.

14.6.2.1 Settlement bank exposures

A settlement bank's maximum intraday exposure to a client can be measured as the client's maximum intraday cumulative net payment (as opposed to receipt) position when it settles directly in the RTGS system. This measure of settlement bank exposure should be regarded as an upper bound because settlement banks can vary the timing of sending clients' payments to minimise their exposure, and require clients to pre-fund settlement obligations.[15]

[15] Note that the timing of settlement in the tiered simulations may also vary depending on the liquidity available to the settlement bank.

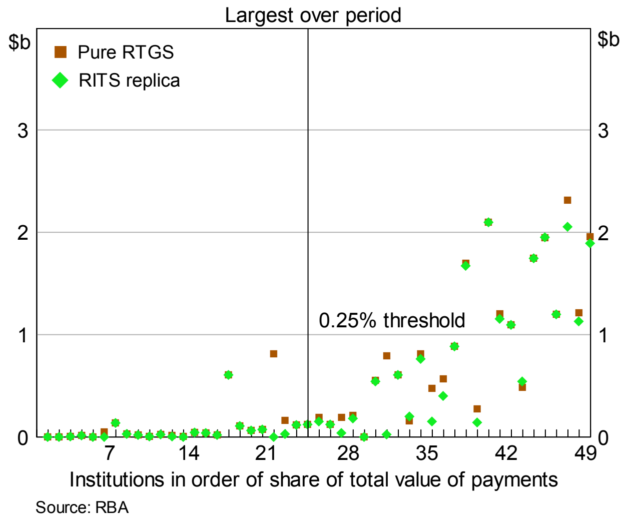
Figure 14.2 **Settlement banks' maximum intraday exposures**



Source: RBA

We find that a settlement banks' *average* maximum intraday exposure to any one of the smallest 29 tiering candidates over the sample period is less than $100 million (Figure 14.2). While the *largest* maximum intraday exposure over the month is roughly three times the size of the average maximum intraday exposure, this is still quite low for the smallest 29 tiering candidates (Figure 14.3). Unsurprisingly, maximum intraday exposures are typically much higher among the largest 20 tiering candidates. We are unable to determine the size of the exposures that the settlement banks in our simulations would be willing to accept, as these are likely to be functions of the capitalisation and risk preferences of the individual institutions. However, we note that while the largest maximum intraday exposure of around $2 billion is of itself sizeable, it is considerably smaller than the tier 1 capital held by each of the four largest settlement banks (over $20 billion in 2008).

Figure 14.3 **Settlement banks' maximum intraday exposures**



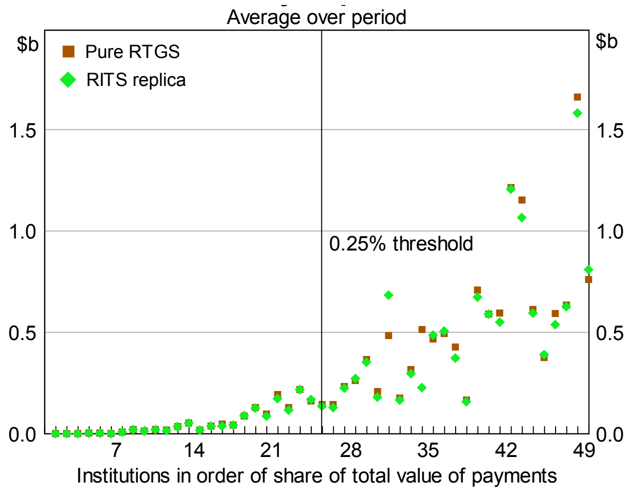
Because our measure of settlement bank exposure (a client's maximum intraday cumulative net payment position) is equal to our measure of the client's liquidity usage when it participated directly in the RTGS system we expect higher settlement bank exposures in the more liquidity intensive pure RTGS system. The difference in exposure between the two system designs varies considerably with the institution being tiered. For the median institution (in terms of this exposure), the average maximum intraday exposure is 8 per cent higher in the pure RTGS system.

14.6.2.2 Individual client exposures

A client's maximum intraday exposure to its settlement bank can be measured using that client's maximum intraday cumulative net *receipt* (as opposed to payment) position when it settles directly in the RTGS system. Because a settlement bank has discretion over the timing of payments, and because it may require pre-funding from its client, these estimates should be viewed as a lower bound.

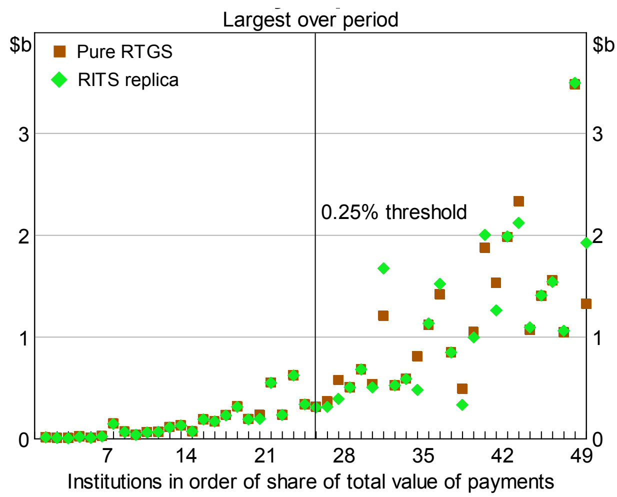
Clients' average maximum intraday exposures are typically less than $1 billion (Figure 14.4). The largest maximum intraday exposures are still less than $1 billion for smaller institutions, but are as high as $3.5 billion for the largest clients (Figure 14.5). Given that the largest clients are typically global banks, their largest exposures are still small relative to their group tier 1 capital.

Figure 14.4 **Clients' maximum intraday exposures**



Source: RBA

Figure 14.5 **Clients' maximum intraday exposures**



Source: RBA

Clients' exposures are typically at least as high in the pure RTGS system as they are in the RITS replica system (although Figure 14.5 shows some exceptions to this among the larger institutions). Again, the percentage difference in exposure between the two system designs varies considerably with the institution being tiered. For the median institution (in terms of this exposure), the average maximum intraday exposure is 2 per cent higher in the pure RTGS system.

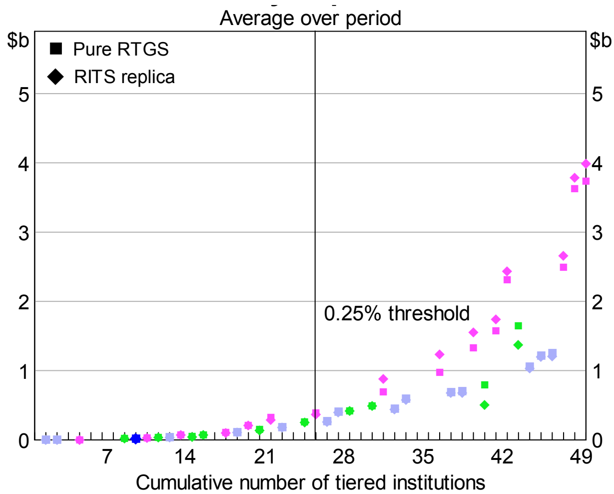
14.6.2.3 Total client exposures

While a settlement bank is unlikely to face the simultaneous default of all of its clients, if a settlement bank defaults all of its clients are exposed. To estimate the maximum total client exposure to a particular settlement bank we can sum the minute-by-minute exposures, measured using each client's cumulative net receipt position when it settled directly.[16] As noted above, these estimates of client exposures should be viewed as lower bounds.

Each observation in Figures 14.6 and 14.7 represents the maximum aggregate loss that could occur if the settlement bank to which the nth smallest institution tiers defaulted on its obligations to those of the n smallest institutions that choose to use it as a settlement bank. For example, when the 49th institution tiers in Figure 14.6, the average maximum intraday exposure in total for that institution and other clients tiered to the same settlement bank is around $4 billion in the RITS replica system.

[16] Note that exposures are not multilaterally netted. Therefore, if a client has negative exposure (that is, it owes the settlement bank), that exposure is excluded from the calculation.

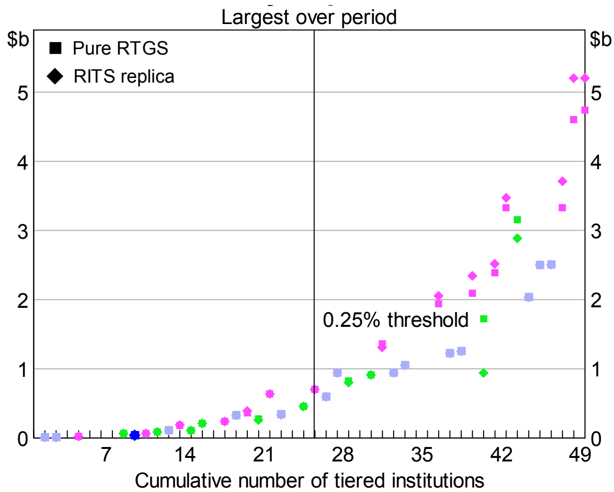
Figure 14.6 **Total client maximum intraday exposures***



* Each colour represents one of the four largest settlement banks. While some tiering candidates initially settle through other institutions, as the simulated level of tiering increases all 49 tiered institutions eventually settle through one of the four largest settlement banks.
Source: RBA

Total client exposures are typically at least as high in the pure RTGS system as they are in the RITS replica system. However, exceptions to this can be seen in Figures 14.6 and 14.7 for cases where larger institutions are tiered to the settlement bank depicted in pink. For the median case, the average maximum intraday exposure is around 1 per cent higher in the pure RTGS system compared with the RITS replica system.

Figure 14.7 **Total client maximum intraday exposures***



* Each colour represents one of the four largest settlement banks. While some tiering candidates initially settle through other institutions, as the simulated level of tiering increases all 49 tiered institutions eventually settle through one of the four largest settlement banks.

Source: RBA

14.7 Weighing the benefits and costs of tiering

Sections 14.5 and 14.6 demonstrate that the liquidity savings from tiering come at the cost of increased concentration and credit risk. It follows that these benefits and costs might be weighed against each other in order to find a socially-optimal level of tiering. This section briefly outlines the considerations and challenges involved in such an exercise. Of course, to do this precisely would require the expression of benefits and costs in measures that are comparable on a dollar-for-dollar basis, which is beyond the scope of this paper.

14.7.1 Measuring the benefit of liquidity savings

One measure of the benefit of liquidity savings is the opportunity cost of the collateral used to obtain liquidity. In the United Kingdom context, James and Willison (2004) suggest that this is equal to the value of the collateral used, multiplied by the spread between the (unsecured) London Interbank Offered Rate (Libor) and the secured-lending repo rate. The intuition behind this calculation is that an institution in possession of collateral-eligible securities could use those securities to obtain funds in the secured lending market, and then lend those funds out at Libor.

In the Australian context, however, there is evidence to suggest that the opportunity cost of collateral is low. The range of collateral accepted by the RBA for intraday repos is significantly broader than that used in secured market trades. Moreover, Commonwealth Government Securities (CGS) are the most commonly-used collateral in intraday repos, and many participants already hold CGS under prudential regulatory requirements. Instead, RBA liaison with RITS participants suggests that the benefit of liquidity savings might be more closely associated with savings in the operational costs (both direct and indirect) of accessing the repo facility. Placing a dollar value on these savings is difficult, given they are likely to vary across institutions.

14.7.2 Measuring the cost of risk

Risk in this context relates to losses that might be realised if a particular event occurs, such as an operational disruption at a participant or the failure of a participant. In general, placing a dollar value on these losses is quite difficult. For example, the incremental social cost of an operational disruption at a settlement bank in a tiered system should take into account the delay and operational costs incurred by:

– the settlement bank itself
– clients of the settlement bank
– other participants in the system
– the operator of the payments system.

Estimating the expected loss due to credit exposures is, in theory, somewhat easier. Section 6.2 provides estimates of the loss that a settlement bank faces if a particular client fails, and vice-versa. Multiplying this potential loss by the relevant probability of failure yields a measure of expected loss, which is comparable on a dollar-for-dollar basis with the benefit of liquidity savings. While probabilities of failure can be inferred roughly from credit ratings, this approach is subject to a number of caveats. Moreover, account should also be taken of the potential second-round effects of the failure of a participant in a tiered system (such as settlement delays and failures at other participants).

14.8 Conclusions

Australia's RTGS system, RITS, has a low level of tiering relative to many RTGS systems elsewhere. This may reflect lower incentives to tier due to particular design features in RITS. The results of the simulations conducted in this paper provide some evidence to support the hypothesis that the design of RITS (that is, an RTGS system with a central queue, a bilateral offset algorithm and a liquidity reservation feature) reduces the incentive to save liquidity by tiering.

While tiering can reduce liquidity needs, it can also increase risk in the system. In terms of credit risk, the simulations provide some evidence to suggest that settlement banks' exposures to clients might be higher in more liquidity intensive systems, although this result is not conclusive. Also, if there were to be an increase in tiering from current low levels, this would result in only small increases to the already high level of concentration in RITS, though it could potentially lead to substantial increases in the share of total payments that individual institutions are responsible for processing.

The results also tend to suggest that both the costs and benefits of the RBA's policy of allowing institutions whose RTGS payments are less than 0.25 per cent of the total value of RTGS payments to tier are modest. Fully quantifying the benefits and costs of tiering to find the socially-optimal level of tiering is left to future consideration. Nevertheless, the results suggest that for institutions below the 0.25 per cent threshold, while settling indirectly only provides modest liquidity savings, it does so without substantially increasing concentration or credit risk. On the other hand, both liquidity savings and risks would increase significantly if institutions above the 0.25 per cent threshold were allowed to tier.

Appendix A

Sub-limits and bilateral offsetting

We have used bilateral limits in the simulator to replicate RITS's sub-limit feature. This involved modifying the simulator's entry and queuing sub-algorithms so that they conduct the appropriate settlement tests (eg test priority payments against a participant's entire settlement account balance, and test active payments against a participant's account balance in excess of its sub-limit). However, data limitations mean that we are unable to pinpoint when a queued payment's status is changed by the sending participant; we only know the status of the payment upon submission to the RITS queue, and the status of the payment when it was settled in RITS. Input to the simulator requires payments to have a single status, which remains unchanged during queuing, thus we had to modify our underlying transaction data. Table A1 summarises the approach.

Table A14.1 **Payment status and submission times**

Status when submitted to RITS	Status when settled in RITS	Status when submitted to the simulator	Time when submitted to the simulator
D	A	A	Settlement time in RITS
	P	P	Settlement time in RITS
A	A	A	Submission time to RITS
	P	P	Submission time to RITS
P	A	P	Submission time to RITS
	P	P	Submission time to RITS

Notes: Payment status: 'D' – deferred; 'A' – active; and 'P' – priority.
In the pure RTGS system design with unlimited liquidity all payments are submitted to the simulator at the time they were settled in RITS and payment statuses are irrelevant.

Payments that were submitted to RITS as deferred are submitted to the simulator at the time that they were settled in RITS. This change is based on the assumption that the sender of a deferred payment did not intend for the payment to settle upon its submission, but rather intended to change the status of the payment at a later time after. (We assume that the actual settlement time in RITS is a better approximation of this time that the time of submission). Payments that were submitted as active but later settled as priority also have their submission time to the simulator changed to their actual settlement time in RITS. A number of participants in RITS have been observed to manage liquidity by setting very high sub-limits, submitting payments to the queue as active, and subsequently changing a payment's status to priority when they want it to be settled. Therefore, again we assume in these cases that actual settlement time in RITS is a better approximation of the time at that the sending participant wished settlement to occur.

We have also designed a bilateral offset sub-algorithm for the simulator that seeks to replicate RITS's own bilateral offset algorithm. In RITS, payments which are queued for over a minute are tested for bilateral offset with up to 10 payments due from the receiving participant on a next-down looping basis.[17] By contrast, the BOBASIC bilateral offset sub-algorithm provided with the simulator only tries to offset all queued payments between the counterparties to the *first* queued transaction, iteratively removing the last queued transaction between these counterparties to find a combination of offsetting transactions that it can settle simultaneously.

[17] We have not incorporated the minute delay feature of RITS's bilateral-offset feature into our sub-algorithm, although this is not expected to affect our results significantly.

Appendix B

Decomposing liquidity savings

To decompose liquidity savings into the two sources identified in the literature, namely liquidity pooling and payments internalisation, we follow Lasaosa and Tudela (2008) and run two additional sets of simulations. For this exercise, we examine the cumulative tiering scenarios.

To isolate the impact of liquidity pooling, we run the tiered simulations including the internalised payments that were previously omitted. This involves transforming payments to and from the client into payments to and from the settlement bank, but continuing to settle payments between the settlement bank and its clients in the RTGS system. Since these internalised payments are still being sent through the system, all the liquidity savings from tiering can be attributed to liquidity pooling.

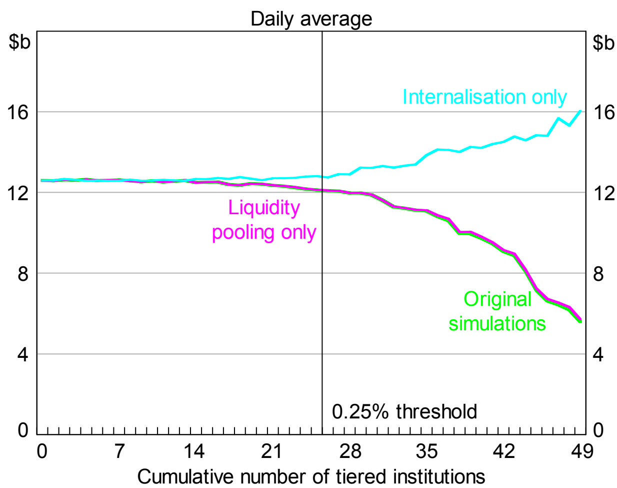
Conversely, to measure liquidity saved due to payments internalisation we omit payments between the client and the settlement bank but otherwise leave the client as a direct participant. Any reduction in liquidity usage in this case will be due to transactions between the client and the settlement bank being settled outside the RTGS system. Note that as multiple clients enter the same tiering network, all payments between them must also be omitted. For example, consider initially that participant B acts as settlement bank for participant A. To measure the internalisation effect when participant B also settles for participant C, payments between participants C, B *and* A must all be omitted.

Since liquidity pooling and internalisation are the only sources of liquidity savings, comparing liquidity savings in the original simulations with those in the additional simulations result in two sets of estimates for the relative importance of the sources of the liquidity savings. Note that these values should be thought of as alternative estimates, not as the upper and lower bound on a range.

There are two reasons for differences in the two sets of estimates for the relative contribution of liquidity pooling and payments internalisation to liquidity savings. First, the complexities of the liquidity recycling process mean that a small change in transaction data can have a substantial effect on the settlement and liquidity profiles. Second, our additional simulations do not perfectly separate out the liquidity-saving effects of tiering. Because the client still participates in the system in the internalisation simulations, the fact that it no longer receives funds from – or pays funds to – the settlement bank creates an artificial and ambiguous effect on its liquidity needs. Note that this effect on liquidity does not exist in the original tiering simulations because in that case the client is completely removed from the system. Hence, this effect could cause the liquidity savings yielded by the internalisation simulations to be materially over- or understated.

Figure A14.1 shows daily average liquidity used in the RITS replica system for different levels of tiering in the original and additional simulations. Note that the green line in this figure is the same as the green RITS replica line in Figure 14.1. Comparing liquidity use in the original simulations to that in the simulations designed to capture the effects of liquidity pooling only suggests that almost all of the liquidity savings from tiering are due to liquidity pooling. Alternatively, comparing liquidity use in the original simulations to that in the simulations designed to capture the effects of internalisation only suggests that well over 100 per cent of liquidity savings are due to liquidity pooling; that is, that the internalisation effect actually *increases* liquidity needs. However, as discussed above, the simulations designed to capture internalisation effects involve an artificial effect on liquidity needs. The results suggest that this effect is putting upward pressure on liquidity needs and thus understating the liquidity saving effect of internalisation.

Figure A14.1 **Liquidity usage decomposition – RITS replica**



Source: RBA

References

Adams, M – Galbiati, M – Giansante, S (2010) **Liquidity Costs and Tiering in Large-value Payment Systems.** Bank of England Working Paper No 399.

Australian Prudential Regulation Authority and Reserve Bank of Australia (2003) **Management of Exchange Settlement Accounts.** Joint Media Release, March.

Bank of England (2009) **The Bank of England's Oversight of Interbank Payment Systems Under the Banking Act 2009.** September 2009.

Bank for International Settlements (2012) **Statistics on Payment, Clearing and Settlement Systems in CPSS Countries.** CPSS Publications No 99.

Chande, N – Lai, A – O'Connor, S (2006) **Credit in a Tiered Payments System.** Bank of Canada Working Paper No. 2006-36.

Chapman J, Chiu, J – Molico, M (2008) **A Model of Tiered Settlement Networks.** Bank of Canada Working Paper No 2008-12.

Ercevik, K – Jackson, J (2007) **Simulating the Impact of a Hybrid Design on the Efficiency of Large-value Payment Systems.** In Leinonen, H (ed), Simulation Analyses and Stress Testing of Payment Networks, proceedings from the Bank of Finland Payment and Settlement System Seminars 2007–2008, Helsinki.

Gallagher, P – Gauntlett, J – Sunner, D (2010) **Real-time Gross Settlement in Australia.** RBA Bulletin, September 2010, 61–70.

Harrison, S – Lasaosa, A – Tudela, M (2005) **Tiering in UK Payment Systems: Credit Risk Implications.** Bank of England Financial Stability Review, December 2005.

Jackson, J – Manning, M (2007) **Central Bank Intraday Collateral Policy and Implications for Tiering in RTGS Payment Systems.** De Nederlandsche Bank Working Paper No 129.

James, K – Willison, M (2004) **Collateral Posting Decisions in CHAPS Sterling.** Financial Stability Review, December 2004.

Kahn, C – Roberds, W (1998) **Payment System Settlement and Bank Incentives.** Review of Financial Studies, 11(4), 845–870.

Kahn, C – Roberds, W (2008) **Payments Settlement: Tiering in Private and Public Systems.** Journal of Money, Credit and Banking, 41(5), 855–884.

Lasaosa, A – Tudela, M (2008) **Risk and Efficiency Gains of a Tiered Structure in Large-value Payments: A Simulation Approach.** Bank of England Working Paper No 337.

Reserve Bank of Australia (1999) **Eligibility for Exchange Settlement Accounts.** Media Release No. 1999-2.

Chapter 15

Liquidity-saving mechanisms: quantifying the benefits in TARGET2

Martin Diehl – Uwe Schollmeyer**

* Martin Diehl and Uwe Schollmeyer are payment system analysts at the Deutsche Bundesbank. This chapter represents their judgements and views and does not necessarily reflect the opinion of the Deutsche Bundesbank.

15 Liquidity-saving mechanisms: quantifying the benefits in TARGET2

Abstract

This paper quantifies the benefits of the liquidity-saving mechanisms (LSM) in TARGET2. It builds on two different models which were developed for the quantification of the benefits of LSM in an environment of fee-based liquidity provision, such as Fedwire, and for a collateral-based payment system. Calibrating with data from 2010 we conclude that considerable positive welfare effects of the implemented LSM in TARGET2 do indeed exist. Depending on the theoretical approach these welfare effects can reach the order of 170.000 to 300.000 Euro per day. However, the institutional setup for the liquidity provision for any specific RTGS has to be taken into account in any case.

15.1 Introduction

The Eurosystem (the ECB and the national central banks of the Eurozone) operates the Trans-European Automated Real-Time Gross Express Transfer System (TARGET2) to ensure the efficient and sound clearing of large-value payments in 23 countries of the European Union. Indeed a smooth functioning of the payment system is a necessary precondition for the functioning of the transmission of the monetary policy and it constitutes the backbone of a resilient market infrastructure which is essential for financial stability. It is therefore obvious that the design of such a large-value payment system is thoroughly and constantly evaluated with regard to both reliability and efficiency.

TARGET2 builds on the user requirements and on the experiences of its predecessor-RTGS of the three providing central banks (Banca d'Italia, Banque de France and Bundesbank). It is equipped with various optimization algorithms as well some different tools for the liquidity management provided to the banks. Since the start of operations in November 2007, TARGET2 is a continuously updated high-end payment system.

The financial benefits of TARGET2 in comparison to a plain-vanilla RTGS have so far never been quantified. A first approach that was conducted by Renault/Pecceu (2007) restricted itself to the increase of the settlement efficiency of a non-FIFO offsetting algorithm compared to a FIFO (first-in first-out) algorithm.

The increased efficiency of a real-time gross settlement system (RTGS) with liquidity-saving mechanisms (LSM) given unaltered behaviour of banks is easy to proof. However, all LSM do also provide an opportunity for strategic withholding of liquidity by single actors. These opportunities may counteract the efficiency gains of the LSM as such. Therefore, models have to be employed which allow for both, liquidity-saving and liquidity-withholding.

The theoretical analysis of liquidity-saving mechanisms (LSM) in large-value payment systems which focused on both aspects starting with Martin and McAndrews (2008) has shown that the introduction of an LSM would normally increase welfare, but under certain conditions welfare might also be reduced. A numerical solution based on simulations presented by Galbiati and Soramäki (2010) gives broadly the same results. In a simulation study based on synthetically created data Schulz (2011) differences between small, medium and large participants and notes that a LSM may have an unequal effect on differently sized banks. When the system is collateral-based instead of fee-based, the introduction of an LSM will arguably always increase the welfare (Jurgilas and Martin, 2010a).

Martin and McAndrews (2008) stated: 'Future research in this area can usefully focus on the question of the empirical magnitudes of the parameters of interest. The important parameters in the model are the cost of delay, the cost of borrowing intraday funds from the CB, the relative size of the payments made to the settlement system versus other payments, and the proportion of time-critical payments. […] [and] the probability that queued payments offset'. Hitherto, the magnitudes of the welfare gains and of the important parameters in the model are still largely unexplored. One sole exception so far is presented in Atalay et al (2010) with regard to the Fedwire System in the USA. This latter paper, however, cannot explore the benefits of a collateral-based payment system as the liquidity provision in Fedwire is fee-based.

Our paper tries to fill these gaps with respect to TARGET2. We show that even a relatively simply modeled LSM saves at about 45.000 to 58.000 Euro per day compared to a plain-vanilla RTGS depending on the equilibrium reached. In a more advanced setup, the savings calculate as about 170.000 to 292.000 Euro per day. This compares to the results of Atalay et al (2010) for Fedwire where the welfare effects are calculated as between $500,000 and $2 million. The different level of savings may in part be explained by the higher transaction value of Fedwire taken into account by the respective authors. In addition, one has to keep in mind that the results are not only depending on such factors as the cost of delay, the cost of the collateral or borrowing intraday funds from the central bank, the relative size of the payments made to the system versus other payments and the proportion of time-critical payments. TARGET2, moreover, features more than a simple LSM. Besides some highly developed queuing arrangements which are the focus of our paper. It also has some other liquidity management tools such as reservations, liquidity pooling and bilateral and multilateral limits. These reservations and limits, however, could potentially also lead to a socially less efficient use of available liquidity, which has to be kept in mind when interpreting the results.

By employing the quoted models of Martin/McAndrews and Jurgilas/Martin we endeavour to support an established line of modeling payment system and behaviour of banks in payment systems. Rather than developing our own model we try to make use of available ideas. Thereby, we hope to contribute to an ongoing process of joint development of payment economics. In addition, in calibrating the models we try to stick as close as possible to the published quantification by Atalay et alii. This contributes to making the results comparable among various large-value payment systems.

The remainder of the paper is structured as follows: In section 15.2 some basic facts about TARGET2 as well as its technical features are presented. Section 15.3 outlines the basic lines of the models developed by Martin and McAndrews (2008) and Jurgilas and Martin (2010a). In Section 15.4 we present our calibrations of the relevant parameters with respect to the conditions in the Eurozone and discuss our findings and compare to the results of other research in this field. Finally in section 15.5 we offer a brief conclusion and propose some field for future research.

15.2 Overview of TARGET2

TARGET2 is an integrated market infrastructure provided by the Eurosystem for the processing of primarily high value and urgent payments in euro. TARGET2 is run by the Eurosystem and is the responsibility of the Governing Council of the ECB. Compared to its forerunner TARGET, which was an association of 17 differential components, TARGET2 is operated on one highly resilient single technical platform (so called Single Shared Platform, SSP). Three Eurosystem central banks – the Banca d'Italia, the Banque de France and the Deutsche Bundesbank (3CBs) – jointly provide this technical infrastructure and operate it on behalf of the Eurosystem. Nevertheless, from a legal point of view, each participating and connected central bank is responsible for the operation of its system component and maintains the business relationships with their local participants.

TARGET2 is Europe's most important payment system for urgent payments and processes a daily average of around 340,000 payments with a total value of almost 2.3 trillion Euro.

In its modular architecture, TARGET2 offers a high degree of flexibility to both central banks and participants. The actual settlement process takes place in the payments module (PM) where each of the 866 direct participants and 69 ancillary systems maintain an account. Intraday liquidity is provided free of interest in the PM either via credit lines on RTGS or central bank accounts (based on a pool of pre-deposited collateral) or via intraday repo transactions with the respective national central bank which is responsible for the business relation with the banks of its country.

Payments can be classified as 'normal', 'urgent', or in exceptional cases as 'highly urgent'. Payments can further be warehoused, ie the submission times can be predetermined. This together with the liquidity management tools determines the payment processing in the entry disposition where a first bilateral optimization mechanism (offsetting of payments) is employed. Normally a basic FIFO mechanism will resolve the payments in the entry disposition. However, in cases where a liquidity increase for (highly) urgent payments would result, normal payments can be processed by a FIFO by-passing principle which is an additional mechanism for saving liquidity.

If the entry disposition fails to settle a payment, this is queued according to its priority status. When in a queue, the settlement manager of a bank can intervene e.g. by reordering transactions within the queue, revoking it or by changing the priority or the set execution time. For the queued payments three different optimization procedures (algorithms) are then available to resolve the queue.

With these optimization procedures and liquidity-saving mechanisms, TARGET2 settles 50% of all transactions within 29 seconds and 90% within 42 seconds. Furthermore, only 0,21% of the volume and 1,8% of the value of all sent payments were not settled on account of a lack of funds or for breaching the sender's limit at the time the system closed.

TARGET2 offers several distinct liquidity management tools for the banks. A direct participant in the payment module has the option to control the use of available liquidity by means of a reservation and a limit system, which may be combined as required. In TARGET2, it is possible for participants to reserve liquidity for urgent and highly urgent payments and to dedicate liquidity to the settlement of ancillary systems. Participants can also define bilateral and multilateral sender limits. Furthermore, banks can use a liquidity pooling functionality within a group to view and use their liquidity, irrespective of the account on which it is held. Increased visibility within the system is also indirectly contributing to more efficient liquidity management. TARGET2 offers online information tools that allow access to all information needed in relation to the payment and liquidity situation of RTGS participants.

Overall, use of liquidity-saving features may depend on several factors. First, it is expected to vary depending on the liquidity situation. Overall use of such features can be expected to be high in tight liquidity situations and low in an environment where liquidity is abundant. Consequently, a relatively low level of recourse to the optimization procedures need not indicate that the liquidity-saving features are inefficient, but that the participants had a sufficient level of liquidity.

15.3 Model setup

15.3.1 Fee-based liquidity provision

Martin and McAndrews (2008) provide for a model of LSM in an environment of fee-based liquidity provision, such as Fedwire. This model is further analyzed for welfare effects by Atalay et al (2008). The model set-up is as follows:

- the day in the payment system is divided into two periods, morning and afternoon,
- participants form a unit mass of banks of equal size but heterogeneous in their payments,
- each bank must make and receive one payment a day,
- a fraction of θ of the banks must make a time-critical payment, for delaying a time-critical payment the banks face delay costs of γ,

- banks may face a liquidity shock in the morning which comes as the net payment to settlement systems;
- a fraction σ receives a positive liquidity shock of size $1-\mu$,
- a fraction σ receives a negative liquidity shock of size $1-\mu$,
- a fraction $1–2\sigma$ receives no liquidity shock,
- banks having a negative balance at the end of the morning must pay an overdraft fee R.

The costs that the banks might face are either R, the overdraft fee from the payment system provider or γ, the delay cost. As it is assumed that there is no market for intraday-liquidity, no interest can be earned by borrowing positive balances through the day.

The provision of a LSM would offer the banks a third option besides borrowing at cost R or delaying at cost γ: queuing. A queued payment will be released when the account of the bank is at least balanced by an incoming payment in the morning period. Implicitly the authors assume the special case of a Balance Reactive Gross Settlement System (BRGS).[1] Finally, all payments are assumed to settle at least in the afternoon period.

The fraction of banks that delay may decrease with ratio γ/R (cost of delay / cost of overdraft). However, the optimal strategy is not so simple: The banks form a belief about the probability of receiving a payment in the morning. And the equilibrium depends on the probability of the liquidity-shock and of the time-critical payments. Martin and McAndrews show, that for some parameter constellations multiple equilibria for both cases (ie with or without LSM) coexist. As the strategy formation involves ex-ante beliefs about the probabilities of receiving a payment in the morning period (π) and the probabilities of being part of the groups of banks that receive a liquidity shock (σ) or a time-critical payment (θ), the socially best solution (the planner's solution) might deviate from the actual solution in the market. In fact, there are fractions λ_i^j of banks with j denoting the banks which delay, queue or pay early and i denoting the membership in the six possible groups defined by the liquidity shock and the time-critical payments.[2]

[1] See Norman (2010) for an overview on different liquidity-saving mechanisms.
[2] See for further details Annex 1 and the original paper.

15.3.2 Collateral-based liquidity provision

Jurgilas and Martin (2010a) extend Martin and McAndrews (2008) to a model for a collateral-based payment system. With the liquidity shock σ of size $1-\mu$ and probability π, the probability for time-critical payment θ and the delay cost γ being equal to the latter model, the cost R is now defined differently. The loss of reputation is assumed to be costly at this rate $R > 0$. Furthermore Jurgilas and Martin have to introduce new variables with regard to the initial level of collateral L_0 posted at the central bank at cost κ per unit. Additional collateral has to be added during the day at cost Ψ per unit. Note that $\Psi > \kappa$. If collateral is added at the end of the day, the cost would be Γ.

The cost function that is to be minimized by the banks is now becoming more complex, as the banks would also have to form a belief ω about the sufficiency of its initial level of collateral (L_0) taking into account the payment activities with other banks denoted Φ. The latter refers to receiving an offsetting payment in the morning period from another bank. The collateral choice is thus crucial for the equilibrium in an environment with and without LSM.

The authors additionally introduce two different cases of payment cycles. Either two payments constitute a short cycle, ie bilateral offsetting, or all payments form a unique long cycle, the latter being deemed more representative for existent payment systems. In short cycles only one equilibrium exists, whereas in a long cycle two equilibria may occur in the case of the absence of a LSM. If a LSM is introduced, there is always exactly one equilibrium and welfare is also always improved compared to a plain-vanilla collateral-based RTGS.

This notion contrasts sharply to the ambiguous result of Martin and McAndrews (2008) regarding the introduction of an LSM into a fee-based payment system. Also the inherent preferences of the banks to delay payments vanish when the liquidity provision is conducted via collateralization. This is so, because the marginal cost of borrowing is zero in the latter case.[3]

[3] See for further details Appendix B and the original paper.

15.4 Quantifying the variables for TARGET2

For the quantification of all variables and calibration of the model with regard to TARGET2 we used information from the national German component of TARGET2 (TARGET2-BBK) for the year 2010. For some variables we counter tested our results with information from interviews with bank managers of larger German banks.

As the deliberations of Atalay et al (2010) had shown, the quantification entails a considerable amount of plausibility considerations since clear construct in the theoretic model do in many cases not match one to one with observable data. Since we try to stick as close to the calibration in Atalay et al (2010) we followed as far as possible their proceedings.

15.4.1 Calibration for a fee-based set-up

Although TARGET2 works with a collateral-based liquidity provision we tried firstly, to assess the welfare effects of LSM in TARGET2 following the fee-based set-up. In doing so we enhanced the comparability of our results towards other calibrations since the first calibration overall was done by Atalay et alii for this sort of set-up. Obviously, some changing assumptions were necessary to get a plausible calibration for TARGET2 and its specific features. They turn out to be quite significant in some cases and are explained for each parameter in detail below.

Calibration of μ and σ

Following Atalay et al (2010) we firstly calibrate the liquidity shock μ and derive a value for the size of the share of banks σ which is subject to this liquidity shock. This can be calculated by the following formula

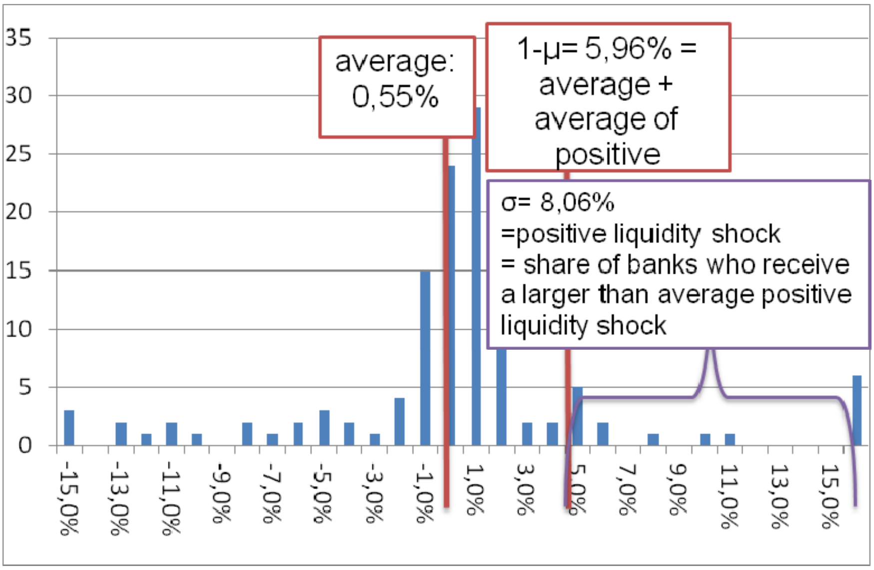
$$1-\mu=\frac{\text{net postion with ancillary systems}}{\text{time}-\text{critical payments}}$$

The average value of net transfers to ancillary systems divided by the time-critical payments is +0.55% in our data.[4] This is the mean value

[4] See for explanation Figure 15.1.

for distribution of banks in TARGET2-BBK. As we concentrate on one side of the distribution, we took the average of the positive values 5.41%. Adding both, gives us the cut-off value $\mu = 5.96\%$ and consequently $1–\mu = 0.94$. As 8.06% of all banks are above this value, we calibrated $\sigma = 0.08$.

Figure 15.1 **Distribution of 1–μ in TARGET2-BBK, 2010**



Source: Own calculations.

However, the distribution of banks according to the share of net-transfers to ancillary systems divided by the time-critical debits in the morning is for TARGET2 not symmetric as was assumed by Atalay et al (2010). If the alternative route via the negative values of the distribution is taken, the values for 1-μ and σ would change to 0.96 and 0.14 respectively.

Calibration of θ, time-critical payments

Next we try to find a plausible value for the size of the fragment of banks with time-critical payments. First of all, the natural approach for the calculation of time-sensitive payments would have been the share of the submitted urgent and highly urgent payments in TARGET2. However, this proved to be inconclusive as it seemed that regarding the present liquidity conditions in the Eurozone as well as the technical efficiency of TARGET2 itself, liquidity managers can be sure that a payment settles in a reasonable time. As was mentioned earlier, more than 90% of all payments settled within one minute, and the share of the entry disposition together with its offsetting mechanism settled around 95% of all submitted payments. In addition, checking with liquidity managers we found out that the use of the 'urgent'- and the 'highly-urgent'-category is extremely heterogenous and does not warrant any conclusions for the true size of time-critical payments. The interviewees gave quite disparate estimations for their time-sensitive payments which in average (7%) were also astonishingly small and they included rather heterogenous sorts of payments different from each other (eg Cash-in-Transit-company payments, securities settlement, money leg of other trades, exceptional third-party transfers). Thus, we discarded this approach early.

We decided rather to follow the approach of Atalay et al (2010) and classified only customer payments as non-time critical (and delete all technical payments) and received $\theta = 0{,}5$, ie a share of 50% time-critical payments in TARGET2-BBK (in terms of value).

Calibration of the delay costs

Calibrating the delay costs (γ) is the next step. Delay costs are not systematically monitored, neither in the US nor in the Eurosystem. Atalay et alii derived them from the model using a technical requirement to hold and calibrating the ratio of γ/R. They could do so, because the overdraft costs are known to the system. TARGET2 does not work with overdraft costs and therefore, we could not apply this model-based calibration. We asked the market participants. They referred to market usance and some institution specific delay costs.

In the case of Clearstream, the delay costs are defined in absolute Euro terms by time brackets and are further differentiated by the cumulative occurrences of delays in a certain time. This is inconclusive for a calculation within the model. A more general approach is the definition of delay costs in the European Interbank Compensation Guidelines of the European Banking Association (Revision 2010). They recommend a value of EONIA added by 0.25 base points plus 100 Euro administration costs. Not regarding the fix part this would in 2010 have amounted to an average of $\gamma = 0.9377\%$.

Calibration of the overdraft fee

As overdraft fees are not applicable to the Eurosystem, and an estimation according to the relation γ/R as in Atalay et al (2010) gave results way off any true collateral cost, we substituted R by κ, the cost of collateral. Not every interviewed bank applied opportunity costs of collateral holding. In addition, the transaction costs of pledging collateral would normally hold a high share of fixed costs. The real costs of pledging collateral differ according to the various sorts of collateral and according to the way of transaction. Thus we calculated for 2010 an average over all classes of collateral of approximately one basis point and applied therefore, $\kappa = 0{,}0001$.

Putting all calibrated data together we calculated the welfare costs for the system without and with LSM using the formula in Appendix A. We concluded that the minimum savings of a simple LSM in TARGET2-BBK that would function on a fee-basis would be 45.000 € per day. If we calculate the second case for the calibration of μ and σ ($1-\mu = 0.96$ and $\sigma = 0.14$), the minimum savings would amount to 58.000 € per day. These numbers are high in comparison to any reasonable calculation of the costs for implementing LSM.

However, these numbers compare to the more impressing 500.000 USD per day from Atalay et al (2010). A significant difference comes from the fact that in Fedwire overdraft costs of six basis points are applied whereas the real collateral costs for TARGET2-BBK are much less (and we substituted the former by the latter for the sake of applicability). Another difference is the total turnover, which was for Fedwire assumed as being much higher than for TARGET2. In addition, the calibration is quite sensitive to the parameters. As has been shown, the model does not entirely fit into the institutional frame of the Eurozone and any additional assumptions would naturally influence the results. We did for some of the reasonable cases also observe, that the absolute values in Euro for the total costs are calculated at negative values, which is an indication for either wrong assumptions or deficiencies of the model employed.

15.4.2 Welfare effects for an inclusion of a LSM within a fee-based liquidity provision

For most of the values of the calibration of the model by Jurgilas and Martin (2008), we could use the above mentioned data. The cost of additional collateral during the day Ψ was assumed to be only slightly higher than the cost for the provision of the level of initial collateral κ because of the general characteristic of the costs (mainly transaction costs) as fixed costs. The same holds true for collateral that is added at the end of the day (Γ). This makes the choice of the initial level of collateral L_0 less crucial. As a consequence we could disregard the belief ω about the sufficiency of its initial level of collateral.

Additionally, some interesting features such as the possibility of auto-collateralization developed by Bundesbank and Clearstream lead to a high amount of collateral available for the purposes of conducting monetary policy with German banks. As Baglioni and Montecini (2008) note, it 'is difficult to provide a reliable estimate of [cost of intraday borrowing from the central bank] because it is not always clear whether a bank is actually constrained to hold those securities or holds them as part of its optimal portfolio management'.

As the model of Jurgilas and Martin (2008) differentiates three possible cases, one of them is not applicable to our setup, we can derive two values for daily savings attributable to the imaginary introduction of a LSM into TARGET2 by using the formula in Appendix B. These welfare effects are calculated as ca. 170.000 € per day and ca. 292.000 € per day, respectively. In comparison to the values for the welfare effects in a fee-based environment, these numbers are higher by the factor four to five and are much closer to the figures for Fedwire by Atalay et alii. Arguably, a set-up within a collateral-based liquidity provision applies better to TARGET2. A minor drawback is just that up to now, we are not aware of any comparable calculations for other RTGS with collateral-based liquidity provision.

15.4.3 Calibration of the necessary collateral

Jurgilas and Martin (2010b) calculate the potential savings in terms of collateral for CHAPS. They conclude that introducing an LSM to CHAPS could reduce the necessary collateral to 50 per cent of the actual level for 2010. Following their reasoning we calculated the level of necessary collateral implied by the model for a collateral-based system and calibrated it with the values as given above. Interestingly, we found that the actual level of collateral used is less than 90 per cent of the minimum suggested by the model in its most favourite case (for an RTGS with LSM). Two explanations occurred to us: Firstly, TARGET2 uses already (for a long time) a sophisticated set of LSM including many other features for managing liquidity as described above. Secondly, the costs for additional collateral are that low that banks do not fear an unsurmountable intraday lack of collateral.

15.5 Conclusion

We applied existing models for measuring the effects of a liquidity saving mechanism (LSM) onto the specific institutional conditions in the Eurozone, namely TARGET2 and its liquidity provisioning mechanisms. We found that even the hardly applicable comparison to a fee-based system such as Fedwire can show that a LSM leads to an increased social welfare in the dimension of about 45.000 to 58.000 € per day. If a better-fitting model with a collateral-based liquidity provision is chosen, the welfare effects are even more pronounced at about 170.000 to 292.000 € per day. The only comparable value from Atalay et al (2010) gives a magnitude of 500.000 USD per day. Both values are comparable since the latter was calibrated according to a much higher turnover.

Both calibrations are not free of reasonable doubts. To apply real numbers to some model values requires in some cases a considerable level of simplification. The adaption of the model to the specific conditions of the Eurozone makes it necessary to estimate some crucial parameters, so that our results can so far only be indicative. Specifically, the share of time-critical payments had to be chosen somehow arbitrary. Further research could follow a number of directions:

– A deeper investigation of the cost of collateral taking into account the heterogeneity of banks.
– An enlargement of the database to all countries of the Eurozone, for the investigation of national structures that have – with regard to payment behaviour – so far not fully integrated despite the multi-national character of TARGET2 and the ongoing financial integration.
– Taking into account the usage of bilateral and multilateral limits by certain banks or national banking communities.
– A refinement of the methodology for investigation of the (marginal) welfare effects of multiple LSM within the same system.
– An improved method for the measurement of some crucial variables such as the time-sensitivity of payments or the (marginal) cost of collateral.

Appendix A

The model of fee-based liquidity provision

Martin and McAndrews differentiate six different type of banks according to two features:

- banks with or without time-sensivite payments (s or r)
- banks with positive, negative or no liquidity shock (s_+, s_-, s_0, r_+, r_-, r_0)

They derive four different equilibria in a world of spontaneous action and three for a social planner. The various types of banks react in each of the equilibria according to the following table, where E stands for ‘sending a payment early’ and D stands for ‘delay’. In the case with LSM the third option ‘Q’ (meaning: queue the payment) occurs.

Equilibria without LSM

Type	s_+	s_0	s_-	r_+	r_0	r_-
1-equilibrium	E	E	E	D	D	D
2-quilibrium	E	E	D	D	D	D
3-equilibrium	E	D	D	D	D	D
4-equilibrium	D	D	D	D	D	D
1-planner	E	E	E	E	E	E
2-planner	E	E	E	E	E	D
3-planner	E	E	D	E	E	D

Equilibria with LSM

Type	s_+	s_0	s_-	r_+	r_0	r_-
1-equilibrium	E	E	E	E	E	E
2-quilibrium	E	E	E	Q	Q	D
3-equilibrium	E	Q	Q	Q	Q	D
4-equilibrium	E	Q	D	Q	Q	D
1-planner	E	E	E	E	E	E
2-planner	E	E	E	E	Q	D
3-planner	E	Q	Q	E	Q	D
4-planner	E	Q	D	E	Q	D

To calculate the welfare effects of an LSM Atalay et al use a calculation of the welfare costs as follows:

W =	Welfare costs
$-\sigma[(\theta\lambda^{e}_{s+} + (1-\theta)\lambda^{e}_{r+})(1-\pi)(2\mu-1)R]$	overdraft costs of banks with positive liquidity shock and who pay early, but did not receive a payment in the morning
$-\sigma\theta\lambda^{q}_{s+}(1-\pi)\gamma$	costs of delaying a time-critical payment of banks who queued, received a positive liquidity shock and did not receive a payment in the morning
$-\sigma\theta\lambda^{d}_{s+}\gamma$	costs of delaying a time-critical payment of banks who delayed and received positive liquidity shock
$-(1-2\sigma)[(\theta\lambda^{e}_{s0} + (1-\theta)\lambda^{e}_{r0})(1-\pi)\mu R]$	overdraft costs of banks without liquidity shock who payed early
$-(1-2\sigma)\theta\lambda^{q}_{s0}(1-\pi)\gamma$	delay costs of banks without liquidity shock who queued
$-(1-2\sigma)\theta\lambda^{d}_{s0}\gamma$	delay costs of banks without liquidity shock who delayed
$-\sigma[(\theta\lambda^{e}_{s-} + (1-\theta)\lambda^{e}_{r-})(1-\mu\pi)R]$	overdraft costs of banks with negative liquidity shock who payed early
$-\sigma[\theta\lambda^{q}_{s-}(1-\pi)\gamma + (\theta\lambda^{q}_{s-} + (1-\theta)\lambda^{q}_{r-}(1-\mu)R]$	overdraft costs of banks with negative liquidity shock who queued
$-\sigma[\theta\lambda^{d}_{s-}\gamma + (\theta\lambda^{d}_{s-} + (1-\theta)\lambda^{d}_{r-})(1-\pi)(1-\mu)R]$	overdraft costs of banks with negative liquidity shock who delayed

Where:

Variable	meaning
σ	fraction of banks with negative liquidity shock = fraction of banks with positive liquidity shock
μ	size of payments between banks
μ	size of liquidity shock
θ	fraction of banks with time-critical payment
γ	cost of delaying time-critical payment
R	overdraft fee
λ^{j}_{i}	fraction of banks that pay early/delay/queue
π	probability of receiving payment in the morning
π^{0}	probability of receiving payment in the morning conditionally on not putting the payment in the queue
π^{q}	probability of receiving a payment in the morning conditionally on putting the payment in the queue

Finally, they multiply the calculated value for W with the turnover of Fedwire and the result is the welfare cost of the respective system.

Appendix B

The model for collateral-based liquidity provision

The sequence of the banks' actions is:

– choose amount of initial collateral: L_0,
– observe liquidity shock λ and liquidity in the morning: $L_1 = L_0 + \lambda(1-\mu)$,
– observe type of payment to be made (time critical or non-time critical),
  – share of time critical payments: θ,
  – delay costs for time critical payments: γ,
– submit a payment (P=1) or delay (P=0) until afternoon,
– with LSM decide if to queue (Q=1) or not (Q=0),
– observe incoming payments,
– post additional collateral at the end of day if needed at costs ψ.

The strategy of the banks is:
– minimize sum of delay and collateral costs
– dependend on liquidity shock, time criticality of payments and on belief about probality to receive a(nother) payment in the morning (ω)

A: The derived solution for an RTGS without LSM is:

$$\min_{L_0} \underset{\lambda,\gamma}{E}\left[\min_{P} \underset{\phi(\omega)}{E}(C_1 + C_2)\right]$$

where:

$$C_1 = \kappa L_0 + PI(L_1 < \mu)(1-\omega^i)(R+\gamma) + (1-P)\gamma$$
$$C_2 = [(1-P)(1-\omega^i) + PI(L_1 < \mu)(1-\omega^i)]\max\{\mu - L_1, 0\}\Gamma$$
$$\Gamma = \frac{(1-\tau_s)^{n-1}}{n}\psi < \psi$$
τ_s: $P = 1$ and $L_1 \geq \mu$
$$L_1 = L_0 + \lambda(1-\mu)$$

There exist multiple equilibria for the choice of the initial level of collateral L_0:

$$\text{if } (1-\mu)\kappa < \gamma\theta(1-\pi) \text{ and } (2\mu-1)\kappa < \gamma\theta(1-\pi): \tag{i}$$
$$L_0 = \mu,\ \omega^i = 1-\pi,\ P^* = 1 \text{ für } \lambda = 0, 1 \text{ und } 0 \text{ fûr } \lambda = -1$$

$$\text{if } (1-\mu)\kappa > \gamma\theta(1-\pi) \text{ and } (3\mu-2)\kappa < \gamma\theta\pi: \tag{ii}$$
$$L_0 = 2\mu-1,\ \omega^i = 1-\pi,\ P^* = 1 \text{ für } \lambda = 1 \text{ und } 0 \text{ fûr } \lambda = -1{,}0$$

$$\text{if } (3\mu-2)\kappa > \gamma\theta\pi \text{ and } (2\mu-1)\kappa > \gamma\theta(1-\pi): \tag{iii}$$
$$L_0 = 1-\mu,\ \omega^i = 0,\ P^* = 0$$

For the case of a social planner two solutions exist:

- if $(3\mu-2)\kappa > \gamma\theta : L_0 = 1-\mu$, whereas $P(\lambda, v, L_0) = 0$ und $P(\mu, \gamma, L_0) = 0,\ \omega^i = 0 \forall \lambda, \gamma$
- otherwise: $L_0 = 2\mu - 1,\ P(\lambda, \gamma, L_0) = 1,\ \omega^i = 1 \forall \lambda, \gamma.$

B: The derived solution for an RTGS with LSM is:

$$\min_{L_0} \mathop{E}_{\lambda,0}\left[\min_{P,Q} \mathop{E}_{\phi(\omega)} (C_1 + C_2)\right]$$

where

$$C_1 = (1-Q)[PI(L_1 < \mu)(1-\omega^i)(R+\gamma) + (1-P)\gamma] + Q(1-P)(1-\omega^q)\gamma + \kappa L_0$$
$$C_2 = \{(1-Q)(1-\omega^i)[(1-P) + PI(L_1 < \mu)] + Q(1-P)(1-\omega^q)\} \text{x} \max(\mu - L_1, 0)\Gamma$$

The optimal collateral choice is the same for the market equilibrium and the social planner:

$L_0 = 1 - \mu,\ P(\lambda, \gamma, L_0) = 0,\ \omega^i = 0 \forall \lambda, \gamma \text{ if } (3\mu - 2)\kappa > \gamma^\theta$

References

Atalay, E – Martin, A – McAndrews, J (2008) **The Welfare Effects of a Liquidity-Saving mechanism.** Federal Reserve Bank of New York Staff Report No. 331. (Originally published 2008, revised 2010).

Atalay, E – Martin, A – McAndrews, J (2010) **Quantifying the Benefits of a Liquidity-Saving mechanism.** Federal Reserve Bank of New York Staff Report No. 447.

Baglioni, A – Montecini, A (2008) **The Intraday Price of Money: Evidence from the e-MID Interbank Market.** Journal of Money, Credit and Banking, Vol. 40(7), 1533–1540.

European Banking Federation (2010) **European Interbank Compensation Guidelines.** Available via https://www.ebaportal.eu/_Download/Legislation/2010/14-06-European_Interbank_Compensation_Guidelines-Revision_2010-D1659K-2009.pdf.

European Central Bank (2011) **TARGET Annual Report 2010.** Available via: http://www.bundesbank.de/download/target2/t2_jahresbericht_2010.en.pdf.

Galbiati, M – Soramäki, K (2010) **Liquidity-saving mechanisms and bank behavior.** Bank of England Working Paper No. 400.

Jurgilas, M – Martin, A (2010a) **Liquidity-saving mechanisms in collateral-based RTGS payment systems.** Bank of England Working Paper No. 389.

Jurgilas, M – Martin, A (2010b) **Liquidity-saving mechanisms in collateral-based RTGS payment systems.** Presentation. Mimeo.

Norman, B (2010) **Liquidity-saving in real-time gross settlement systems – an overview.** Bank of England Financial Stability Paper No. 7.

Martin, A – McAndrews, J (2008) **Liquidity-saving mechanisms.** Journal of Monetary Economics, Vol. 55(3), 554–567.

Renault, F – Pecceu, J-B (2007) **From PNS to TARGET2: the cost of FIFO in RTGS payment systems.** In Leinonen, H (ed.) Simulation studies of liquidity needs, risks and efficiency in payment networks. Bank of Finland Studies E:39.

Schulz, C (2011) **Liquidity Requirements and Payment Delays Participant Type Dependent Preferences.** European Central Bank Working Paper No. 1291.

TARGET2-SSP General Functional Specifications (Version 2.1). Available via: http://www.bundesbank.de/download/target2/target_2_user_v2_1.pdf.

Chapter 16

The Mexican experience in how the settlement of large payments is performed in the presence of high volume of small payments*

*Biliana Alexandrova-Kabadjova** – Francisco Solís-Robleda****

* Preliminary version – please do not quote

** Payment Systems, Banco de Mexico, E-Mail: balexandrova@banxico.org.mx.
*** Payment Systems, Banco de Mexico, E-Mail: fsolis@banxico.org.mx.

16 The Mexican experience in how the settlement of large payments is performed in the presence of high volume of small payments

Abstract

Payment systems play a key role in the financial infrastructure of all modern economies. Participants of payment and other settlement systems need access to intraday liquidity to fulfill their payment obligations. They do that either using their own funds, which are costly or recycling incoming payment. In order to relay on incoming payments, banks could delay the settlement of their own payment obligations. From the regulators' point of view is important to know to what degree participants relay on the payments they receive from others. In Mexico, this is among the first studies that analyze the intraday liquidity management of the Mexican Real Time Settlement Payment System, SPEI from this perspective. We examine a data set of transactions in order to get insights of the participants' behavior regarding the delay of sending payment orders. To that end, we use transactional data from April 7 to May 7, 2010. All payment instructions that arrived from 9:00 a.m. to 17:00 p.m. each working day are included. For the study we use an artificially created environment that reproduces SPEI's operational conditions.

16.1 Introduction

Payment systems have evolved over time as modern economies are becoming more and more dependent on the services they provide. Their key role in the financial infrastructure is changing from being used to discharge large financial market payment obligations, to becoming an important service provider not only for all kinds of businesses, but also for individuals. The technological innovation and the increased awareness of saving cost by using electronic payments are among the main drivers for changes in the payment service industry. Ongoing innovation is likely to diversify even more payment types competing on consumer service level, whereas efficiency and cost reduction could be the main reasons for integration of payments processing and settlement. The development of those processes with opposite directions, diversification and integration, differ across countries. For instance, in the European Union, despite the specific country's requirements and preferences of payment services, the need for international integration has driven the creation of two payment platforms – the cross border financial infrastructure, the Single Euro Payments Area (SEPA) and the Large Value Payment System (LVPS) TARGET2, which nowadays are among the most advance examples of standardization and process integrations (Kokkola, 2010).

In this context, it is worth to highlight that the volume of direct credit transactions has experienced an important growth in the last years. In Canada, for instance the volume of direct credit transactions has rose from 857.3 million in 2005 to 1,201 million in 2010, overcoming the annual volume of cheques transactions since 2009 and nowadays is the second instrument in terms of relative importance, after payment cards, according to BIS (2011a) and BIS (2011b). In another example, in the Euro Area the use of credit transactions is also growing, even though this payment method there is the third in relative importance. It has rose from 12,391 million reported for 2001 to 16,187 million reported form 2010 according to European Central Bank. In Mexico, even though in a different scale the volume, electronic retail payments has grown significantly. In 2002 there were 884 million non-cash transactions including cheques, card payments (credit and debit cards) and electronic transactions (direct credit and direct debit), whereas in 2010 these number rose 1.6 times to 2,300 million transactions.[1]

This tendency could have two possible consequences, which need to be considered. The first is the growing demand for urgent small payments and the second is the increased aggregated value of the transactions, which with the time could become systemically important. Adding to those two factors the necessity of cost reduction, it could be that in the near future, real time high-value payments and low value electronic payments may be settled together. To achieve this, due to the volume of transactions settlement engines need to ensure highly efficient liquidity usage with the guaranty that that retail payments do not delay time sensitive payment that settle important financial market obligations. In this line of research, mainly the efficient use of liquidity, several studies have been developed in the last decades among which are Armantier, Arnold and McAndrews

[1] Source: Central Bank of Mexico.

(2008), Becher, Galbiati and Tudela (2008) and Denbee and Norman (2010). To that end, payment systems need to establish timely and liquidity–efficient operational rules, which will allow the settlement of a high volume of retail payments with a minimum pressure on intraday liquidity usage. This issue has become in the last decade central bank policy concern, not only because central bank usual are operators of the LVPS, but also from the perspective of the regulator Ball, Denbee, Manning and Wetherilt (2011), Johnson, McAndrews and Soramäki (2004), Leinonen and Soramäki (2001) and Manning, Nier and Schanz (2009).

SIX Interbank Clearing SIC (the Swiss Interbank Clearing System), with volume of 394.7 million transactions reported for 2010, is the best known example of a LVPS that processes retail payments BIS (2011b). Also, other countries like the Czech Republic, Serbia, Slovakia, Turkey, Ukraine and Mexico use one system to settle wholesale payments and low value payments together Allsopp, Summers and Veale (2009). In Mexico, an important rate of low value direct credit transactions between banks go through a Real Time Settlement Payment System, SPEI, together with the settlement of large value payments. The system is operated by the Mexican Central Bank and it settles payment orders on real time, charging its participants a 0.50MXN per payment. In the year 2010 processed 80.1 per cent of the volume of the Large Value Payments in the country, according to BIS (2011b). SPEI processes, on average, around 500,000 operations daily. More than 80% of the transactions are payments with value lower than 100,000 MXN and, only 1.3% of the transactions are above 10,000,000 MXN.

One of the most important advantages in using real-time gross settlement (RTGS) systems for settling Large Value payments is the elimination of settlement and credit risk that could arise between participants (further referred also as banks) in deferred net settlement systems (Bech, 2008). Nevertheless, as a consequence, RTGS systems require relatively large amounts of intraday liquidity to support payment obligations, in comparison with deferred net settlement systems. This liquidity can be sourced from the participant's funds, usually in the form of intraday overdrafts obtained from the RTGS operator (the central bank) or from incoming payments from other participants. In that way, by delaying payments the banks' behavior determines the underling structure of the payment instructions' settlement.

In the present study our interest is focused on the participants' decision to delay payment orders, with the assumption that an individual liquidity usage is determined by the size and urgency of the payment transactions (Norman, 2010). Nevertheless if urgency is not strictly demanded, the delay in the settlement of payment orders could reduce the level of individual liquidity usage. Here, it is important to clarify that if banks do not know in advance the size and the time of incoming payments (which is assumption we have), the delay of payments could be a signal that the participant would like to relay on the incoming payments to settle her own obligations (Bech, 2008, Galbiati and Soramäki, 2011, and McAndrews and Potter, 2001). However, given the complex interdependence game among banks and the operational rules in place in the particular payment systems, delaying payment orders no necessarily implies less liquidity usage per participant. Furthermore an imbalanced interdependency among banks could raise concerns about the level of settlement risk triggered by individual behavior, which eventually could turn out to have systemic consequences. In this context operational rules could play a crucial role to avoid that participants depend excessively on the incoming payments for settling their own obligations. Nevertheless before a proper framework to evaluate the participants' behavior is established, we believe that better understanding of the banks' intraday liquidity management is required and in particular we are interested to know which factors are taken into account in the delay of payments.

To that end, in the present paper, in order to get further insights of the motivation behind the participants' decision to delay payments, we analyze a set of payment orders of SPEI. Continuing with the line of study presented in Alexandrova-Kabadjova and Solis-Robleda (2012), in which the authors analyze from a more general perspective the intraday liquidity management, here our aim is to study how the delay of settling payment obligations is related to the different size and volume of the transactions. In particular given that SPEI settles together wholesale and retail payments, we have the opportunity to evaluate the delay of payment, when settling in real time a large number of low-value payment transactions. We use payment transactions from April 7 to May 7, 2010 sent to SPEI from 9:00 a.m. to 17:00 p.m. The data for each payment transaction includes payer, payee, amount of the transaction, time of reception and time of settlement. The currency used is Mexican Pesos MXN.[2] For our analysis, we perform two cases of study – the first is an empirical study and it is presented in two parts, the second is performed in

[2] 1MXN = 0.076USD or 1MXN = 0.057EUR according to www.xe.com on 26 of April 2012.

simulation environment. The rest of the paper is organized as follows – in the next section we briefly present the notation used for our study, following in section 16.4 the results obtained in the first part of our empirical study are shown. Afterwards in section 16.5 we present the second part of the empirical study and in section 16.6 the result obtained from the simulation study are exposed. Finally, in section 16.7 we conclude with our mayor findings and suggestions for future research.

16.2 Structure and notation

In this section we describe the variables used to measure the settlement delay. For each case of study we measure the delay differently, but before we explain how we do that, let us start with a brief exposition of how the study is organized together with some of the operational rules of SPEI.

For the empirical study we divide the transactional data in four subsets, according to the value of the payment orders – first subset contains the transactions with value lower than 100[3] MXN, second subset includes payments with value between 100 and 1000 and the transactions with value equal to 100, third subset contains payment requests with value between 1000 and 10000 and transactions with value equal to 1000 and finally the fourth subset includes payment orders with value higher than or equal to 10 000.

In the first part of this study we evaluate the impact of the payments delayed on a level of intraday liquidity usage. We start by showing the general picture of the transactional data, ie on a daily bases we present per each subset of payment instructions a histogram of transaction, a time structure of all payments and a time structure of delayed payments; average time of delay per payment, the aggregated time of delay, the average amount of per transaction and the aggregated level of value of payment orders. . In the second part of the empirical study we focused more specifically on the delayed payments. Per each subset of transaction previously defined, we calculate the daily average time of delay, the daily proportion of payment instructions delayed and the daily average amount of transactions. In order to determine the difference among the four subsets, we compare the statistical measures obtained.

[3] From now on all references to the value of the payment transactions are in thousand, at least other way stated.

In the second case of study in order to evaluate the overall impact of the postponed payments, we define a specific measure [4] ν, which represent on the aggregated level millions of MXN per minute delayed. For that reason we divide the data in four subsets, but this time in the following way. The first subset contains all payments, the second include transactions with values higher than or equal to 100 MXN, the third contains payment orders with value higher than or equal to 1000 MXN and finally the fourth subset include transactions with value higher than or equal to 10000 MXN. In a simulation-based environment reproducing the operational conditions of SPEI, we process each subset of transactions separately. We then make a comparison between the first set (all payments) with each one of the rest of the subsets. In order to compare the calculated delay ν more accurately, we consider for our analysis always the same set of transactions. Thus from the first subset (all payments) only those higher value payments, which are included in the compared subset are present in the benchmark value. In other words, for the set of all payment three times we calculate ν – for the transactions with value higher than 100 MXN, for the payment orders with value higher than 1000 and finally for the payment instructions with value higher than 10000. We compare the specific ν obtained for each subset with the ν corresponding of the first subset.

Regarding the operational rules of SPEI, the system receives payment instructions continuously during the day, which are placed in a queue. It closes operations at 17:35 and starts processing payments for the next working day at 19:00. During operation time, a settlement process (SP) is executed at the latest 20 seconds after receiving a new payment. Payment instructions, which are not settled in a certain SP are kept in the queue and are considered for settlement in the subsequent processes. After execution of the latest SP before the operation is closed, payments in the queue are cancelled.

The intraday liquidity needed to support payment obligations, can be sourced from the participant's funds, usually in the form of intraday overdrafts obtained from the RTGS operator (the central bank) or from incoming payments from other participants. The later allows banks to recycle liquidity in order to offset outgoing payment instructions (Bech, 2008, McAndrews and Martin, 2008, and Norman, 2010). That way the liquidity cost of making payments is reduced, as participants avoid incurring overdrafts from the central bank, which

[4] For details of how we calculate the delay in millions of MXN per minute, see equation (16.1).

require pledging collateral or maintaining high quality securities (government debt) for repos. In other words, the amount of liquidity used depends among other factors on the time of sending payment orders and on the particular sequence and size of the transactions, which are strategic decisions of the participants (for more detailed analysis of the factors determining the liquidity usage see 19). In that way, in order to efficiently use the different source of liquidity, the intraday liquidity management consists of a careful scheduling the settlement of payment orders throughout the day (Bech, 2008).

Nevertheless the information revealed from the transactional data we use does not reflect for how long the payments sent to SPEI have been at the participants' own queue. We can only observe from the data the difference between the reception and settlement time of each payment. According to the operational rule of SPEI, the reason why a payment is not settled in the next SP after reception is because there are either not available funds in the sender's account to cover, neither there are incoming payments to offset the payment request.

In order to measure the settlement delay, let P is the set of payments received and processed by SPEI in one day and I is the set of participants in SPEI such that $p_{i,j}$ is a payment in P from the participant $i \in I$ to participant $j \in I$; $t_{p_{i,j}}$ is the sum of minutes passed from the first SP launched immediately after the reception of the payment instruction $p_{i,j}$, until its settlement;[5] $\varphi_{p_{i,j}}$ is the amount in MXN of $p_{i,j}$. The settlement process is performed either every 20 seconds or after the reception of 300 payments, wherever happens earlier. For the propose of our analysis, we consider that regardless the reason of delay the payment is delayed if $t_{p_{i,j}} > 1\min$, which means that there were at least two or more settlement processes before the payment finally was settled.

In our second case of study, the aggregated delay ν of settlement during one day, measured in terms of billions of Mexican pesos in one minute, is defined on the following way

$$\nu = \sum_{p_{i,j} \in P} \left(t_{p_{i,j}} \varphi_{p_{i,j}} \right) \qquad (16.1)$$

[5] Note that $t_{p_{i,j}}$ could represent a fraction of a minute.

It is worth to highlight that if the funds of the participants are sufficient to settle every $p_{i,j} \in P$ in the next SP launched immediately after the reception $p_{i,j}$ then $\nu = 0$.

In order to perform this test we need to define the minimum required level of intraday liquidity. In Large Value Payment Systems, the term intraday liquidity is used to define the funds that the participants have to cover their payment obligations during one day. Those funds come primary from two sources: firstly participants' resources from previous balances or electronic transactions from other payment systems and secondly from payments received during the day from the rest of the participants. For our study, we establish the minimum required level of liquidity in terms of participants' resources. To that end, we have defined several measures listed on what follows. First, for each $i \in I$ the intraday payment orders sent are presented as

$$P_{snt} = \sum_{j \in I} p_{i,j} \tag{16.2}$$

whereas the received payments are denoted as

$$P_{rcv} = \sum_{j \in I} p_{j,i} \tag{16.3}$$

We define l_i^{min} as the lower required level of liquidity for settlement during the day, which is determined as follows

$$l_i^{min} = \max\{(P_{snt} - p_{rcv}), 0\} \tag{16.4}$$

16.3 How volume and value is related to the delay of payments

In this section we present the histogram and the time structure of the volume of transactions per each subset previously defined as well as the corresponding time structure of the delayed payments.

With the aim to distinguish specific characteristics of banks' strategies related to the way participants are sending payment requests to SPEI, on what follows we list the behavior patterns observed per each subset of payments. For this analysis we have organized the data according to the value of the transactions, as previously defined – first subset contains the transactions with value lower than 100 MXN, second subset includes payments with value between 100 and 1000 and the transactions with value equal to 100, third subset contains payment requests with value between 1000 and 10000 and transactions with value equal to 1000 and finally the fourth subset includes payment orders with value higher than or equal to 10 000. In all figures in this and in the next sections the studied period is presented with 23 different gray tone used for each day. We have used the same color representing the same day on all figures, starting with 7th of April 2010 (Wednesday) and ending with 7th of May 2010 (Friday). The size of the transactions subsets per day are presented in table 16.1.

Table 16.1 **Number of transactions per subset**

Calendar day	**Week day**	**Lower than 100**	**Between 100 and 1000**	**Between 1000 and 10000**	**Higher than 10000**
07.04.2010	Wednesday	155,105	22,028	4,980	2,136
08.04.2010	Thursday	152,486	23,330	5,245	2,318
09.04.2010	Friday	224,792	30,200	5,951	2,608
12.04.2010	Monday	157,765	22,302	5,798	2,758
13.04.2010	Tuesday	137,454	21,011	5,563	2,577
14.04.2010	Wednesday	182,802	25,552	5,897	2,322
15.04.2010	Thursday	230,521	27,196	5,955	2,526
16.04.2010	Friday	250,805	31,548	6,255	2,621
19.04.2010	Monday	156,275	24,359	6,382	2,981
20.04.2010	Tuesday	139,684	19,808	4,805	2,999
21.04.2010	Wednesday	134,356	20,428	4,831	2,633
22.04.2010	Thursday	145,622	22,934	5,826	3,215
23.04.2010	Friday	220,872	30,596	6,069	2,792
26.04.2010	Monday	151,798	22,890	5,699	3,061
27.04.2010	Tuesday	138,812	21,867	5,375	2,645
28.04.2010	Wednesday	149,549	24,625	5,870	2,692
29.04.2010	Thursday	210,685	30,810	6,802	3,572
30.04.2010	Friday	340,610	42,498	8,830	4,023
03.05.2010	Monday	151,932	20,546	5,794	2,807
04.05.2010	Tuesday	142,973	19,613	5,129	2,823
05.05.2010	Wednesday	164,877	27,077	7,009	2,836
06.05.2010	Thursday	162,148	22,994	5,659	3,047
07.05.2010	Friday	233,578	29,760	6,417	3,665

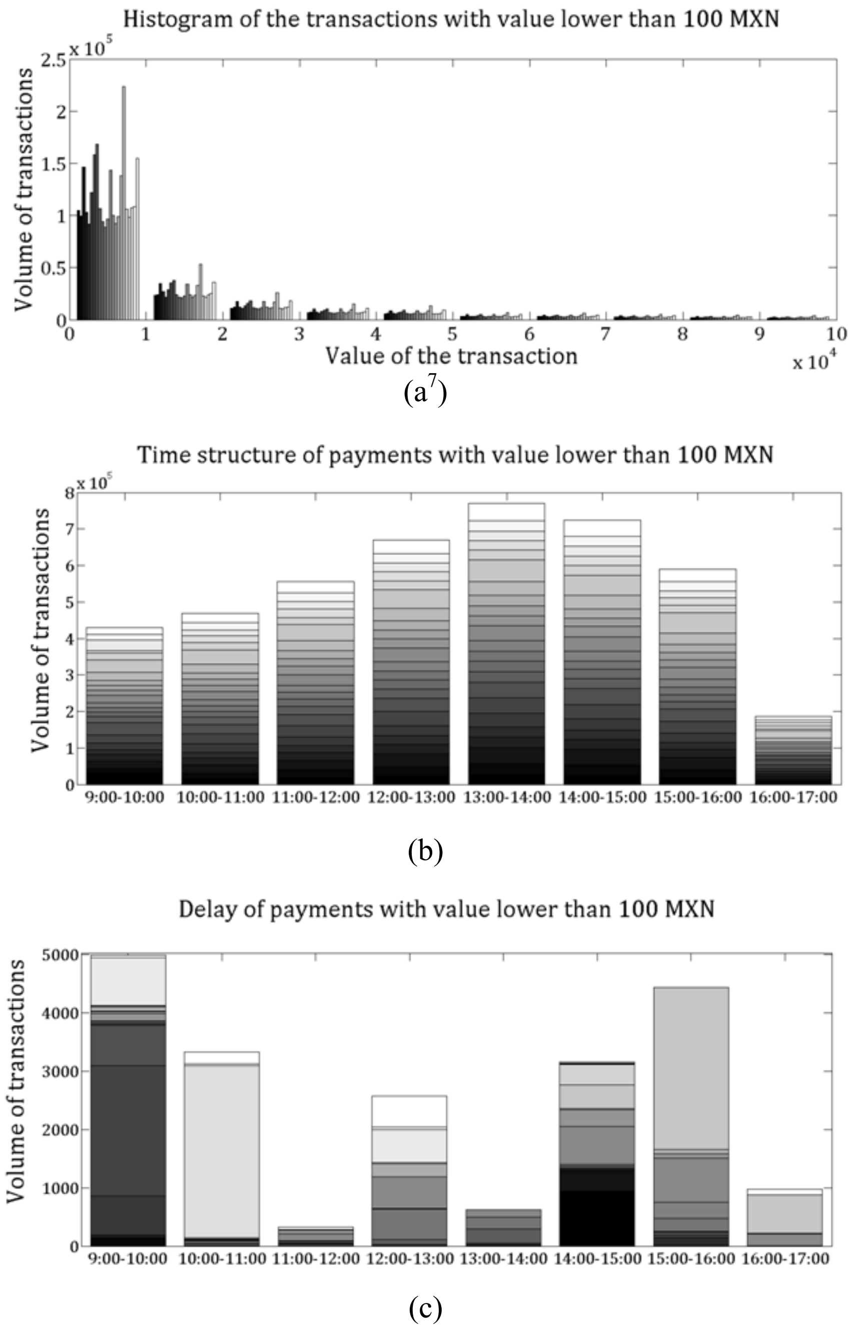
In figure 16.1 we present the subfigures elaborated with transactions having values lower than 100 MXN. In particular in subfigure (a) the histogram of the number of payment instructions is shown, in subfigure (b) we present the time structure of the transactions on aggregated level and in subfigure (c) on the same way the time structure of the number of payments delayed is presented.

We observe at the subfigure 16.1(a) that in normal operational conditions there are certain patterns in the way participants sent payment orders with value lower than 100. In particular, given the average and standard deviation calculated per day of the week presented in the first two rows in table 16.3, we can say that there is a weekly periodicity in such a way that Fridays the transaction volume is higher than the rest of the week, whereas Tuesday is the day with lower volume.[6] In comparison with the daily average and standard deviation presented in table 16.3, the week day average calculated for Monday, Tuesday and Wednesday is lower than the overall daily average presented in table 16.3, whereas the standard deviation observed per each one of the days of the week presented in table 16.2 is smaller than the standard deviation reported in table 16.3. Furthermore, Monday and Wednesday present the same average. In addition, the peak in terms of transaction volume observed in subfigure 16.1(a) corresponds to the last working day of the month, which is considerably higher than the rest. Finally we notice that the volume of low value transaction presented in the first decile of the histogram is above 100,000 transactions daily, which is significantly bigger than the rest of the histogram's deciles.

With respect to the time structure of payment with value below 100 presented in subfigure 16.1(b), other regularity is observed. The volume of low value payments reported per hour is considerable throughout the day. The hour with highest volume of transactions is between 13:00 and 14:00, followed by the transactions sent between 14:00 and 15:00. This observation could be an indication that participants prefer to send low value payment during the afternoon hours.

[6] The 23 days of our simple includes 4 weeks and 3 days, starting at Wednesday and ending at Friday. The last working day of the month is Friday on the penultimate week of our example.

Figure 16.1 **The case of transactions with value lower than 100 MXN**



[7] In all figures the value of payment transactions are presented in thousands of MXN.

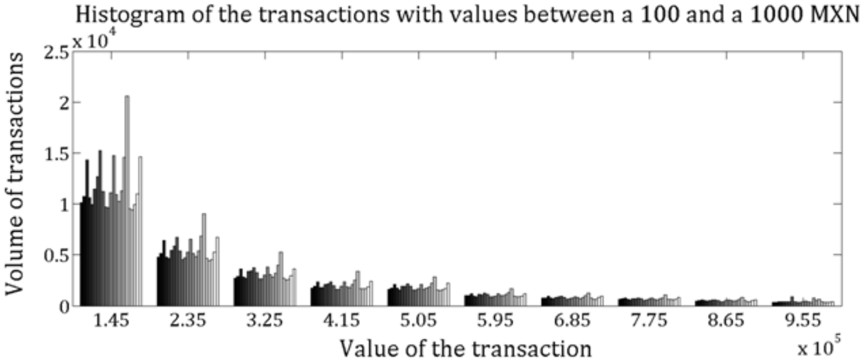
Table 16.2 **Week day average and standard deviation per subset**

		Monday	Tuesday	Wednesday	Thursday	Friday
<100	Avg	166,164	150,471	166,134	192,061	267,119
	Std Dev	2,691	2,063	15,789	34,316	46,366
100 – 1000	Avg	25,426	23,336	25,179	28,388	36,151
	Std Dev	1,447	796	2,154	3,339	5,218
1000 – 10000	Avg	5,931	5,218	5,307	5,897	6,704
	Std Dev	266	284	474	512	1,075
>10000	Avg	2,902	2,761	2,385	2,936	3,142
	Std Dev	124	164	237	457	588

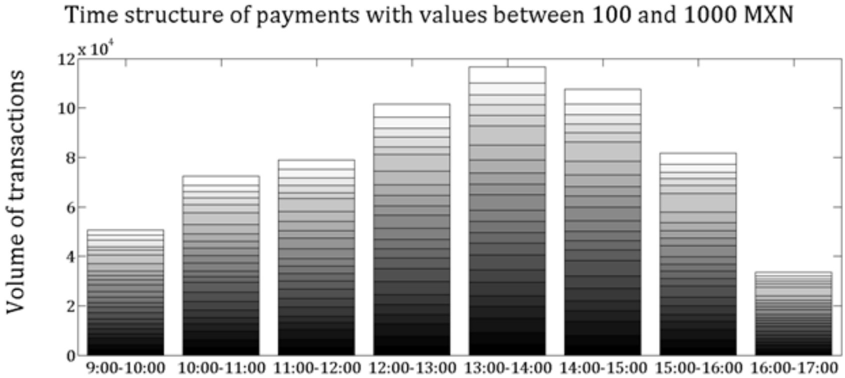
Regarding the time structure of the payments delayed shown in subfigure 16.1(c), a regularity on a daily bases is not observed. We also notice that on certain days the number of delayed payments is significantly higher than others. According to the total and hour correlation coefficient between the number of transactions and the number of payments delayed presented in table 16.4, we can say that for this subset of transactions the hour with higher number of payments delayed is between 15:00–16:00 and 16:00–17:00 followed by 9:00 and 10:00, whereas a negative correlation is observed for the hours between 10:00–11:00 and between 13:00 and 14:00. Finally we observe in subfigure 16.1(b) that the highest volume of payment transaction is between 13:00 and 14:00.

Next, let us look at the Figure 16.2, in which the transactions with value between 100 and 1000 MXN are presented. In particular, as in the previous case in subfigure 16.2(a) the histogram of the number of payment orders is shown. In this subfigure a similar weekly regularity is observed, which is consistent with the week day average and standard deviation presented in table 16.2, in which Tuesday is the day with the lower level of transactions and Friday with the highest. Furthermore, Monday and Wednesday again have the same average as listed in table 16.2. Nevertheless, for this subset the difference among deciles is not that strongly underlined as in the histogram of the lowest payments.

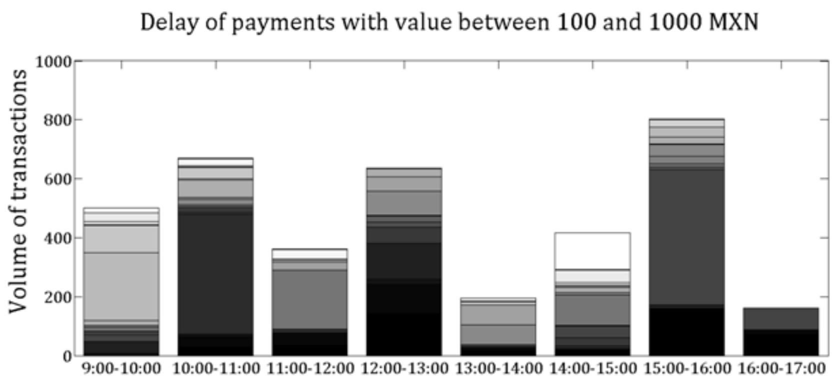
Figure 16.2 **The case of the transactions with value between 100 and 1000 MXN**



(a)



(b)



(c)

Table 16.3 **Daily average and standard deviation per subset**

	Daily Average	Standard Deviation
< 100	191,004.87	50,662.94
100–1000	27,984.43	5,564.87
1000–10000	5,832.22	822.53
> 10000	2,824.57	456.43

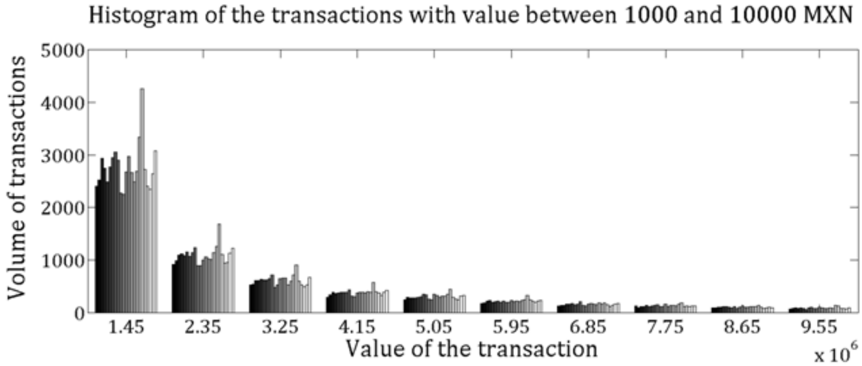
Regarding the time structure of the payment transactions presented in subfigure 16.2(b), a regularity per hour is clearly observed, with the highest volume of transactions reported at 13:00–14:00 and then at 14:00–15:00. We notice that in comparison with subfigure 16.1(b), here the volume observed between 9:00 and 11:00 is lower than the volume of transactions from the rest of the day before 16:00 o'clock.

Table 16.4 **Coefficient of correlation between number of delayed payments and total transactions per subset**

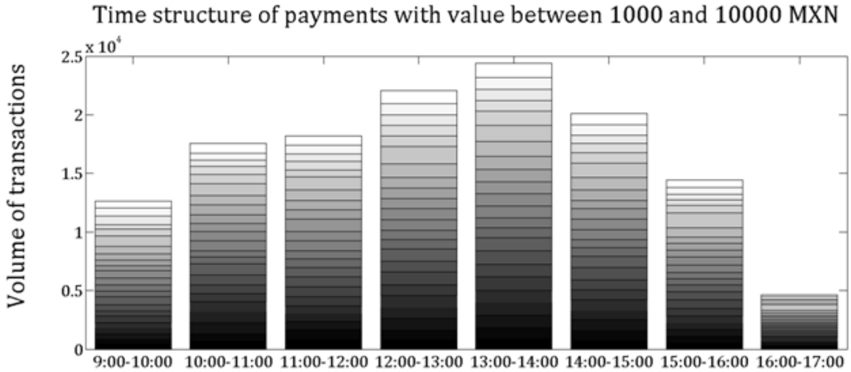
	< 100	100–1000	1000–10000	> 10000
Total	0.183	0.171	0.114	0.639
9:00–10:00	0.314	0.039	0.107	0.741
10:00–11:00	–0.040	–0.092	–0.110	0.705
11:00–12:00	0.231	0.280	0.369	0.739
12:00–13:00	0.046	0.016	0.045	0.332
13:00–14:00	–0.090	–0.090	–0.002	0.303
14:00–15:00	0.198	0.037	–0.109	0.118
15:00–16:00	0.775	0.838	0.675	0.790
16:00–17:00	0.828	0.778	0.874	0.623

With respect to the hourly structure of the delayed payments shown in subfigure 16.2(c), we observe that across days and hours the number of delayed payment varies. We focus our analysis at the hours before 16:00 o'clock and we notice that between 15:00 and 16:00 is the hour with the large number of payment delay, which also present the highest correlation coefficient listed in table 16.4. In that table, for this subset of transactions a negative correlation is observed again between 10:00 and 11:00 and between 13:00 and 14:00. In addition, in this case the higher volume of payment orders observed at subfigure 16.2(b) is between 13:00 and 14:00 o'clock.

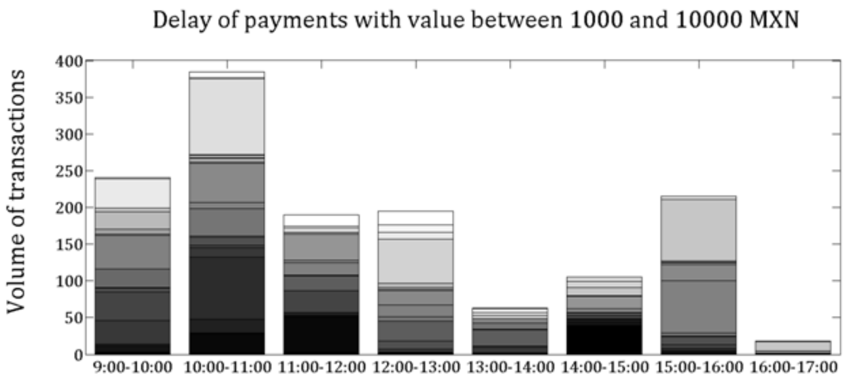
Figure 16.3 **Transactions with value between 1000 and 10000 MXN**



(a)



(b)



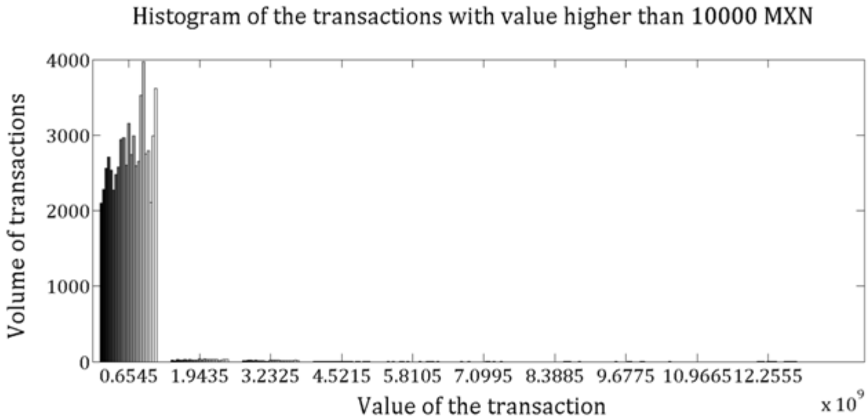
(c)

Following, in figure 16.3 we present the subfigures corresponding to the subset of transactions with value between 1000 and 10000 MXN. In this case the picture observed changes with comparison to the previos two figures. Accourding to the data presented in table 16.2, the differences among the week day average is not that substantial, with the exception presented on Fridays, in which the highest volume of transactions is reported.

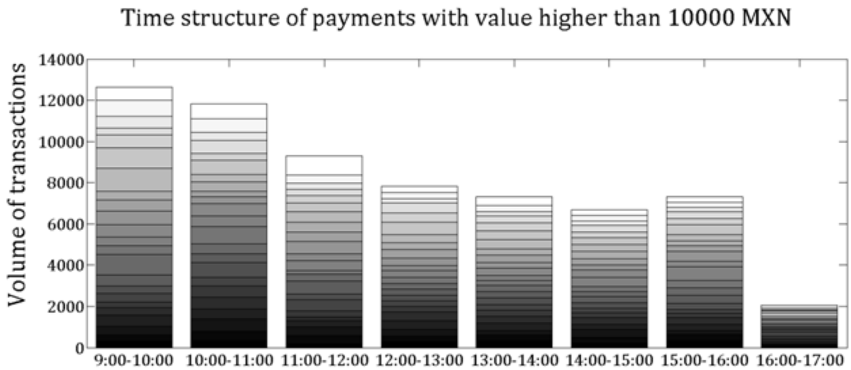
Nevertheless, in subfigure 16.3(b) the pattern in the time structure of the transactions is similar to the one observed in subfigure 16.2(b), but in this case the highest vulume of the payment orders is observed between 13:00–14:00 and between 12:00–13:00 and the daily volume is steady. Regarding subfigure 16.3(c), in which the time structure of delayed payments is reported, differently form the previously discripted subset of transactions, here the delay of payments is observed on a daily base. Eventhough, the total correlation coefficient presented in table 16.4 is the lowest among the analysed four subset. Furthermore, for this case during the day we observed three hours with negative correlation coefficients. In the previos two cases, the negative correlation are observed between 10:00 and 11:00 and between 13:00–14:00, here in addition to those, a negative correlation is also observe between 14:00 and 15:00 o'clock.

Finally the graphs obtained from the subset of transaction higher than 10 000 are presented in figure 16.4. For this case in subfigure 16.4(a) we observe a different histogram with comparison to the previous subsets. In this subfigure, the majority of payments are concentrated in the first decile and the way payments are sent throughout the week do not follow a pattern. This observation is supported by the week day average presented in table 16.2, in which no significant difference is observed among the average of the days of the week. Following with our analysis, in subfigure 16.4(b) we notice that the time structure of sending payments is different from the previous three cases. We observe a clear participants' tendency to send large value payments during the morning operational hours. Furthermore as shown in subfigure 16.4(c) and in table 16.2 the number of delay payments is correlated with the volume of payment sent. This high level of correlation (above 0.70) is not presented in the subsets of transactions with lower values. Consequently, the total correlation coefficient for the payment transactions with value higher than 10000 MXN, is the highest. We can also say that the delay of payments is performed on a daily bases, but there is no regular pattern in terms of number of payments per day. Among hours before 16:00 o'clock, the hour with the lowest number of payments delay is between 13:00 and 14:00, but in this case the correlation is not negative, as in the previous three cases.

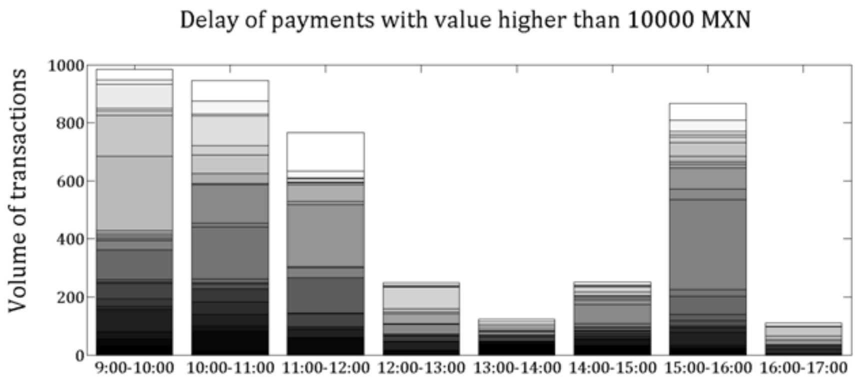
Figure 16.4 **Transactions with value higher than 10000 MXN**



(a)



(b)



(c)

In order to finalize the observation made in this section, based on our analysis we can conclude that the way payments are send to SPEI depend on the value of the requiest. In general we observe that the participants' behavior is not random with the highest volume of transactions observed on Fridays and the lowest on Tuesday. Looking at the volume of transacitions throughout the different days of the week, according to the data presented in table 16.2, we can say that payments with value lower than 1000 are following similar patterns, whereas the volume of payment orders with value higher than 1000 are steady among different days.

On the other hand, the intraday data included in this section allows us to study the way payments are sent throughout the day, for which we can visualize three patterns:

Payments with value lower that 1000 MXN. The highest volume processed of those payment orders is observed between 13:00 and 15:00 (please refer to figures 16.1(b) and 16.2(b)). In addition, those transactions do not have in general a high correlatation between the number of payment delayed and the number of payments processed (it is around 0.17). Evenmore, this corelation is negative for the hours 10:00 to 11:00 and 13:00 and 14:00. Finally, during the day the highest values of this correlation is observed between 15:00 and 17:00 hours (please refer to table 16.4).

Payments with value between 1000 and 10000 MXN. In this case the highest volume of transactions processed is oberved between 12:00 and 14:00 (please see figure 16.3(b)). Furthermore, this set of transactions has in general the lower correlation between delayed payments and number of payments processed. During the day, this correlation is negative between 10:00 and 11:00, 13:00 and 14:00 and also between 14:00 and 15:00, whereas the highest level is observed between 15:00 and 17:00 hours as presented in table 16.4.

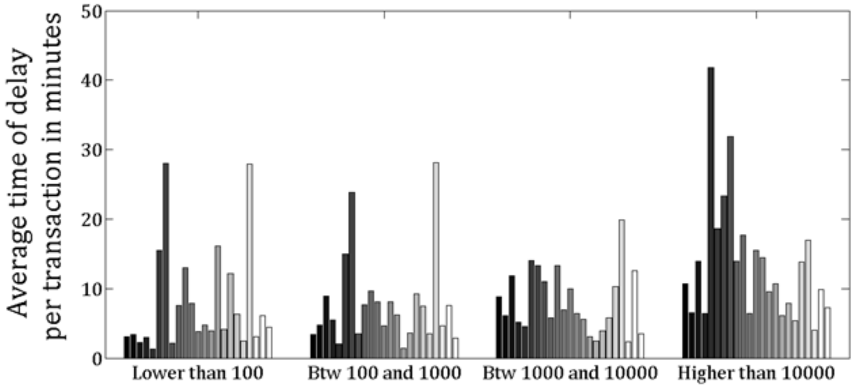
Payments with value higher than 10000 MXN. The highest volume of those transactions is processed between 9:00 and 12:00 o'clock (please refer to figure 16.4(b)). Further, this set in general has the highest correlation between delayed payments and number of payments processed and it is possitive throughout the day. In particular between 9:00 and 12:00 the value of this correlation is above 0.7 and it is high also between 15:00 and 17:00, as we notice in table 16.4.

16.4 Measuring the delay in terms of time, volume and amount

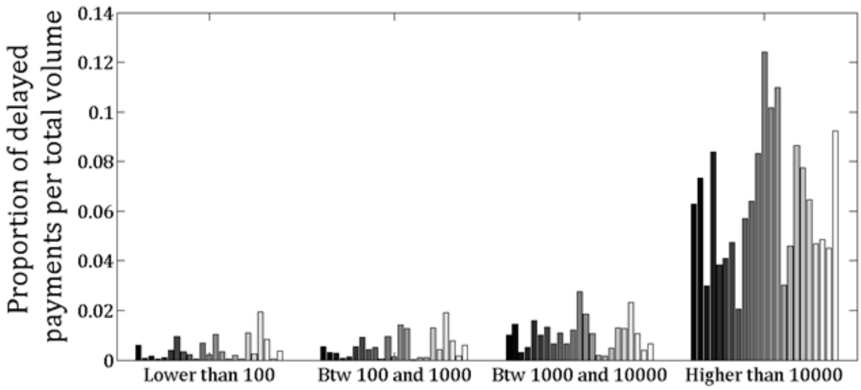
In order to go further in our analysis, we present in this section three additional measures of the delayed payments calculated per each one of the four subsets of transactions defined in the previous section. The measures are calculated on a daily bases and are the following – the average time of delay per payment order, the proportion of payments delayed per volume of the subset transactions, and the average amount per payments delayed. The results of those calculations are presented in figure 16.5.

In subfigure 16.5(a) we present the average time of delay per transaction calculated according each subset, whereas in table 16.5 some general statistics are listed regarding this measure. We observe in table 16.5 that the statistics reported for payment orders with value lower than 100 and between 100 and 1000 have very similar characteristics, in particular the average, the minimum and the maximum. Regarding the transactions with value between 1000 and 10000, we can say that the maximum time of delay is the lowest among the four subsets. This observation goes with the same line presented in the previous section, in which we said that this subset of transactions report the lowest number of payments delayed. Finally regarding the transactions with value higher than 10000, we observe that those payments on average spend more time waiting to be settled and the minimum and maximum time of delay is considerable higher than the rest of the transactions.

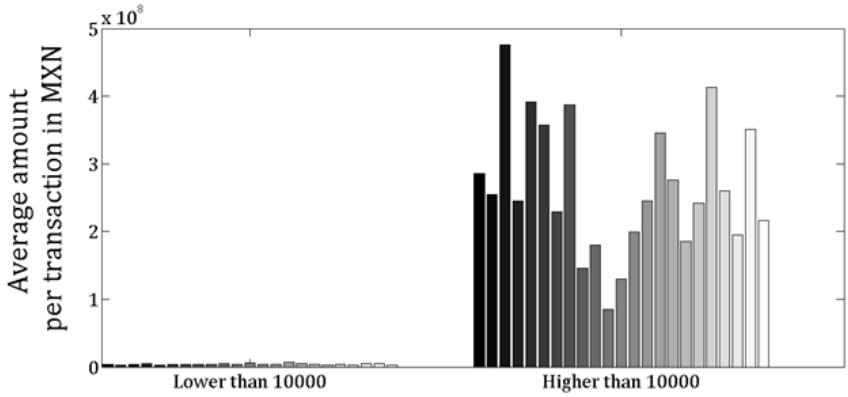
Figure 16.5 **Measuring different aspects of the delayed payments**



(a)



(b)



(c)

Table 16.5 **Statistics on the average time of delay**

	>100	**100–1000**	**1000–10000**	**<10000**
Average	7.93	7.81	8.12	13.60
Median	4.48	6.17	6.41	10.71
Std. Dev.	7.63	6.54	4.56	8.98
Min	1.32	1.42	2.34	4.05
Max	28.03	28.16	19.91	41.83

Following in subfigure 16.5(b) we present the proportion of the number of payments delayed in comparison to the total volume of transactions. In table 16.6 and 16.7 are shown the general statistics regarding to the number of payments delayed and the proportion of the number of payment delayed, respectively. We observe in subfigure 16.5(b) that the patters of payment instructions with value lower than 100 and instructions with value between 100 and 1000 are very similar. With respect to the subset of transactions with value between 1000 and 10000, we notice an increase in the proportion of payments delayed, but the share is still considerable lower than the proportion of payment orders delayed in the subset of transactions with value higher than 10000. This subfigure and the data presented in table 16.7 clearly show that payments with value lower than 1000 follow similar behavior rules presenting the lowest proportion of delayed payments (on average 0.0043 and 0.0056 respectively). In addition we observed that the proportion of delayed payments doubled for the subset of transactions with values between 1000 and 10000 (on average 0.0106), and increased significantly – six times – for the payments with value higher than 10000 (on average 0.0641).

Table 16.6 **Statistics on the number of payments delayed**

	>100	100–1000	1000–10000	<10000
Average	901	199	65	183
Median	387	142	59	170
Std. Dev.	1043	232	41	102
Min	53	6	9	50
Max	3867	959	160	399

Table 16.7 **Statistics on the proportion of the number of payments delayed**

	>100	100–1000	1000–10000	<10000
Average	0.0043	0.0056	0.0106	0.0641
Median	0.0024	0.0042	0.0105	0.0627
Std. Dev.	0.0047	0.0052	0.0066	0.0273
Min	0.0004	0.0002	0.0015	0.0206
Max	0.0194	0.0192	0.0275	0.1241

Finally we look at the average amount per transactions, which in subfigure 16.5(c) is presented only as a division between payment orders with value higher than 10000 and those orders with value lower than 10000. In table 16.8 we present the statistical measures for the four subsets of transactions, which are shown as the total aggregated value of transactions. This data suggests that delayed payment orders of large value could have a significant impact on the liquidity usage through the day.

Table 16.8 **Statistics regarding the average amount of delayed payments**

	>100	**100–1000**	**1000–10000**	**<10000**
Average	16,803.54	364,795.27	3,620,199.51	264,973,104
Median	16,711.15	338,263.46	3,353,865.1	245,633,951
Std. Dev.	4,172.35	83,749.86	1,005,163.05	98,412,065.3
Min	9,026.01	273,173.03	2,346,977.92	85,208,203.5
Max	24,759.30	609,629.35	6,671,626.71	475,550,938

16.5 Stressing the intraday liquidity

In this section we present the results of a simulation test aim to evaluate the impact of low value payments on the settlement of large value payments, given that intraday liquidity is limited. In order to reproduce the operational conditions of SPEI, we use an artificial environment. We elaborate the simulation scenarios with the same 23 days transactional data, which are structured in four sets, delimited according to their value: set one include all payments; set two is a subset of payments with value higher than 100 MXN; set three is a subset of payments with value higher than 1000 MXN and finally set four is a subset of payments with value higher than 10000 MXN. In table 16.9 we present the size of the subsets per day. We measure the effect of settling in real time a large volume of low-value payment by calculating the settlement delay ν of the large value per each subset previously defined.

Table 16.9 **Number of transactions per subset**

Day	**All Payments**	**Payments >= 100**	**Payments >= 1000**	**Payments >= 10000**
07.04.2010	195,637	31,280	7,116	2,136
08.04.2010	195,578	33,211	7,563	2,318
09.04.2010	277,326	41,367	8,559	2,608
12.04.2010	202,695	33,616	8,556	2,758
13.04.2010	179,899	31,728	8,140	2,577
14.04.2010	229,436	36,093	8,219	2,322
15.04.2010	279,731	38,203	8,481	2,526
16.04.2010	305,347	43,045	8,876	2,621
19.04.2010	205,322	36,703	9,363	2,981
20.04.2010	181,098	30,611	7,804	2,999
21.04.2010	174,978	30,525	7,464	2,633
22.04.2010	193,068	35,190	9,041	3,215
23.04.2010	274,774	42,249	8,861	2,792
26.04.2010	198,330	34,711	8,760	3,061
27.04.2010	182,009	32,532	8,020	2,645
28.04.2010	196,682	35,879	8,562	2,692
29.04.2010	269,387	44,756	10,374	3,572
30.04.2010	416,860	59,374	12,853	4,023
03.05.2010	195,294	31,954	8,601	2,807
04.05.2010	184,136	30,388	7,952	2,823
05.05.2010	217,316	39,758	9,845	2,836
06.05.2010	208,648	34,747	8,706	3,047
07.05.2010	290,832	43,507	10,082	3,665

In order to perform the liquidity stress test, first, according to equation 16.4 we calculate for each of the subsets above the minimum required level of participants' funds, l_i^{min} for which all transactions are settled for each subset. Here we assume that the settlement order of the transactions will not be modified, neither the number of payments will be reduced by the changes in the intraday liquidity level. This is a very strong assumption, as other studies have proven that under conditions of stress, participants' behavior changes and not only the order the payments are send could change, but depending on how sever the stress conditions are the most probable scenario could be to reduce significantly the volume of transactions (McAndrews and Martin, 2008). Nevertheless, given that we do not have insights which would be the reaction of the participants in SPEI, we decide not to modify the volume of transactions, neither the order of payments. Nevertheless due to the reduction of the available liquidity, the underline structure analyzed in the previous sections is no longer the same.

Table 16.10 **Statistics on the average settlement delay**

		Payments >=100		Payments > = 1000		Payments >10000	
		Settled with all payments	**Settled separately**	**Settled with all payments**	**Settled separately**	**Settled with all payments**	**Settled separately**
ν	Average	11,487	11,636	11,179	11,198	11,138	11,100
	Std. Div	3,331	3,328	3,497	3,592	3,493	3,321

For our study, we use the simulator with transactions corresponding to each one of specified subsets and calculate the settlement delay ν for subsets two, three and four. Then we compare each one of them to the settlement delay ν of the first subset. In order to make the comparison more accurate, we include in each of the three cases only the payments corresponding to the transactions with higher value. Thus from the subset of all payments only the transactions with higher value are included, which correspond to the transactions of the compared subset. We present the comparison in the settlement delay ν in figure 16.6, 16.7 and 16.8 respectively. In addition in table 16.10 we present the average and the standard deviation of ν for the three cases of study in billions of MXN.

We observed form the figures 16.6, 16.7 and 16.8 and the average presented in table 16.10 that settlement delay ν is not significantly modified by the inclusion of low value payments. We notice that the sequence and the size of payments in a particular day are factors that determined more significantly the millions of MXN per minute delayed than the division by value for the subsets of transactions.

Figure 16.6 **The settlement delay ν in conditions of stress (first comparison)**

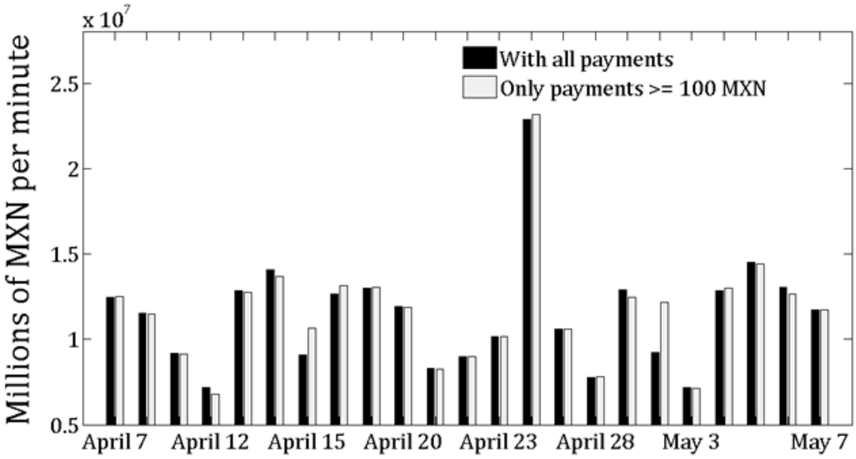


Figure 16.7 **The settlement delay ν in conditions of stress (second comparison)**

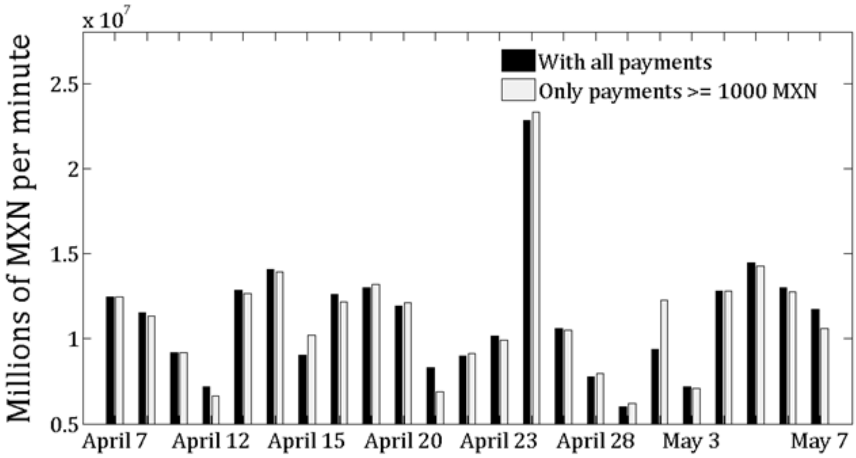
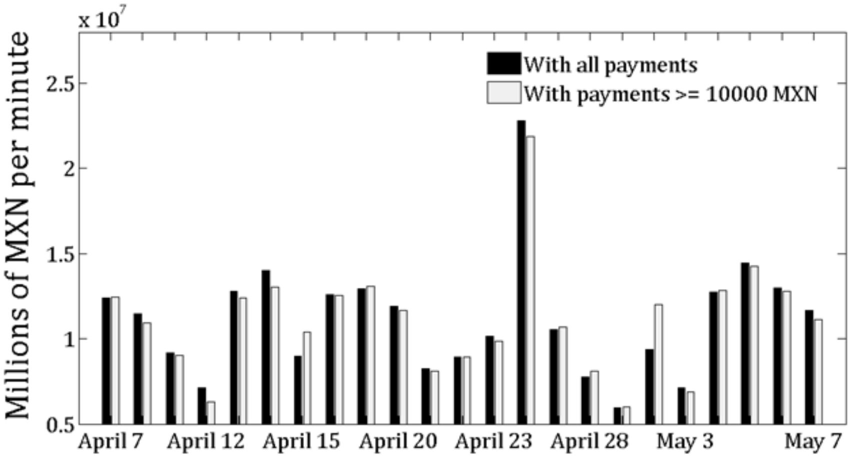


Figure 16.8 **The settlement delay ν in conditions of stress (third comparison)**



16.6 Conclusions

This is among the first studies of intraday liquidity analysis of the Mexican Large Value Payment System SPEI. This paper looks at the number of payments delayed from two perspectives. The first objective is to get further insights regarding the motivation of the participants in SPEI to delay payments in operational conditions and the other, is to evaluate the impact of low value payments in the settlement of large value payment in condition of stress.

To that end we elaborate two cases of study using thirty days transactional data taken from SPEI between 7 of April and 7 of May 2010 corresponding to the payment orders performed from 9:00 a.m. to 17:00 p.m. each working day. Both cases are performed in an artificially created environment that reproduces the operational conditions of SPEI. The first case allows us to make observation on the emerging patterns of participants' behavior given the empirical evidence by dividing the transactions in four subset determined by their value. In particular, for each one of these subsets we look at three aspects of the payment orders – the histogram, the time structure of transactions, and the time structure the number of payments delayed. We also include per subset a week day average and standard deviation in table 16.2 and daily average and standard deviation in table 16.3. Furthermore, we calculate a correlation coefficient between a number of payments delayed and a number of payments processed. Those coefficients are presented in table 16.4.

What we can conclude from our observations is that a clear weekly regularity is observed in the volume of payments with value lower than 1000 MXN. With respect to the volume of large payments this patter is not observed. Further, regarding the time structure thought the day for sending payments, we observe that it follows different patterns for large and low value payments. In particular we can divide the transactions in three categories – payments lower than 1000, transactions between 1000 and 10000 and payment request higher than 10000. Moreover, the majority of large value payments are sent during morning operational hours, in which a high correlation coefficient between delay payments and number of payments processed, is observed. On the other hand low value payments have a peak observed between 13:00 and 14:00, which is negatively correlated with the number of payments delay during this hour.

The second case of study has allowed us to evaluate the settlement delay of millions of MXN per minute delay ν measured in conditions of limited intraday liquidity. According to the results, we have observed that low value payments do not increase the settlement delay. Furthermore, what determines the level of ν are the sequence and the size of payment orders per particular day, primary determined by the large value payments.

As a final remark, we believe that more studies related to the intraday liquidity management are required in order to get further insights of the participants' behavior. One possible extension to the present work could be to analyze the intraday liquidity usage in relation to the observed delayed payments. We also could apply the empirical analysis to a more extensive period of time, which statistically will be more accurate.

References

Alexandrova-Kabadjova, B – Solis-Robleda, F (2012) **Managing intraday liquidity. The Mexican Experience.** In Simulation in Computational Finance and Economics: Tools and emerging Applications, eds. Alexandrova-Kabadjova, Martinez-Jaramillo, Garcia-Almanza and Tsang, IGI Global, 143–160.

Allsopp, P – Summers, B – Veale, J (2009) **The evolution of realtime gross settlement: access, liquidity and credit, and pricing.** Financial infrastructure series, Word Bank.

Armantier, O – Arnold, J – McAndrews, J (2008) **Changes in the timing distribution of Fedwire Funds transfers.** Economic Policy Revie, Federal Reserve Bank of New York, 14: 83–112.

Ball, A – Denbee, E – Manning, M – Wetherilt, A (2011) **Intraday liquidity: risk and regulation.** Bank of England Financial Stability Paper, 11:1–25.

Bech, M (2008) **Intraday liquidity management: a tale of games that banks play.** Economic Policy Review, Federal Reserve Bank of New York, 14:7–23.

Bech, M – Garratt, R (2003) **The intraday liquidity management game.** Journal of Economic Theory, 109:198–219.

Becher, C – Galbiati, M – Tudela, M (2008) **The timing and funding of chaps sterling payments.** Economic Policy Review, Federal Reserve Bank of New York, 14:113–133.

BIS (2011a) **Statistics on payment and settlement systems in selected countries – figures for 2009.** Report 95, Bank for International Settlements.

BIS (2011b) **Statistics on payment and settlement systems in selected countries – figures for 2010.** Report 98, Bank for International Settlements.

Denbee, E – Norman, B (2010) **The impact of payment splitting on liquidity requirements in RTGS.** Bank of England Working Paper No. 404, 2010.

European Central Bank. Statistics for payment instruments and large-value and retail payment systems. Figures for 2010. Retrieved from http://www.ecb.int/stats/payments/paym/html/index.en.html.

Galbiati, M – Soramaki, K (2011) **An agent-based model of payment systems.** Journal of Economic Dynamics and Control, 35:859–875.

Johnson, K – McAndrews, J – Soramaki, K. (2004) Economising on liquidity with deferred settlement mechanisms. Economic Policy Review Federal Reserve Bank of New York, 10:51–72.

Kokkola, T (2010) **The Payment System.** European Central Bank Publications.

Leinonen, H – Soramaki, K (2001) **Optimizing liquidity usage and settlement speed in payment systems.**

Manning, M – Nier, E – Schanz, J (2009) **The economics of large-value payments and settlement.** Oxford University Press.

McAndrews, J – Martin, A (2008) **Liquidity savings mechanisms.** Journal of Monetary Economics, 55:554–567.

McAndrews, J – Potter, S (2002) **Liquidity Effects of the Events of September 11, 2001.** Economic Policy Review, Federal Reserve Bank of New York, 8:59–79.

McAndrews, J – Trundle, J (2001) **New payment system designs: cause and consequences.** Bank of England Financial Stability Review, 127–136.

Norman, B (2010) **Liquidity saving in real-time gross settlement systems – an overview.** Bank of England Financial Stability Paper, 7:1–11.

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.
hgfhf

E:43 Hanna Freystätter **Essays on small open economy macroecomics**. 2012. 169 p. ISBN 978-952-462-793-1, print; ISBN 978-952-462-794-8, online.

E:44 Vesa Ronkainen **Stochastic modeling of financing longevity risk in pension insurance**. 2012. 124 p. ISBN 978-952-462-801-3, print; ISBN 978-952-462-802-0, online.

E:45 Matti Hellqvist and Tatu Laine (eds.) **Diagnostics for the financial markets – computational studies of payment system. Simulator Seminar Proceedings 2009–2011.** 2012. 458 p. ISBN 978-952-462-813-6, print; ISBN 978-952-462-814-3, online.

ISBN 978-952-462-813-6
ISSN-L 1798-1077

Edita Prima Oy
Helsinki 2012