WORKING paper

BANQUE DE FRANCE
EUROSYSTÈME

Payments delay: propagation and punishment

Ben Craig[1], Dilyara Salakhova[2] & Martin Saldias[3]

March 2018, WP # 671

Abstract

We use a unique dataset of transactions from the real-time gross settlement system TARGET2 to analyze the behavior of banks with respect to the settlement of interbank claims. We focus on the time that passes between a payment's introduction to the system and its settlement, the so-called payment delay. Delays represent the means by which some participants could free ride on the liquidity of others. These delays are important in that they can propagate other delays, thus prompting concerns that they could cause system gridlock. This paper characterizes the delays in the TARGET2 and analyzes whether delays in incoming transactions could cause delays in outgoing transactions. We distinguish between the potentially mechanical pass-through of delays and the reaction of one bank to its delaying counterparty, and we propose a set of instruments to tackle endogeneity issues. We find evidence that delays do propagate downstream; however, in most cases the effect is rather limited. As for delaying strategies on a payment-by-payment basis, contrary to the theoretical literature, the data show only very weak evidence. This conclusion opens a venue for research how banks may rather follow persistent liquidity management routines.

Keywords: Payment delays, endogenous regressors, liquidity, TARGET2

JEL classification: C26, E42, G21

[1] Federal Reserve Bank of Cleveland, Deutsche Bundesbank, ben.r.craig@clev.frb.org

[2] Banque de France, dilyara.salakhova@banque-france.fr

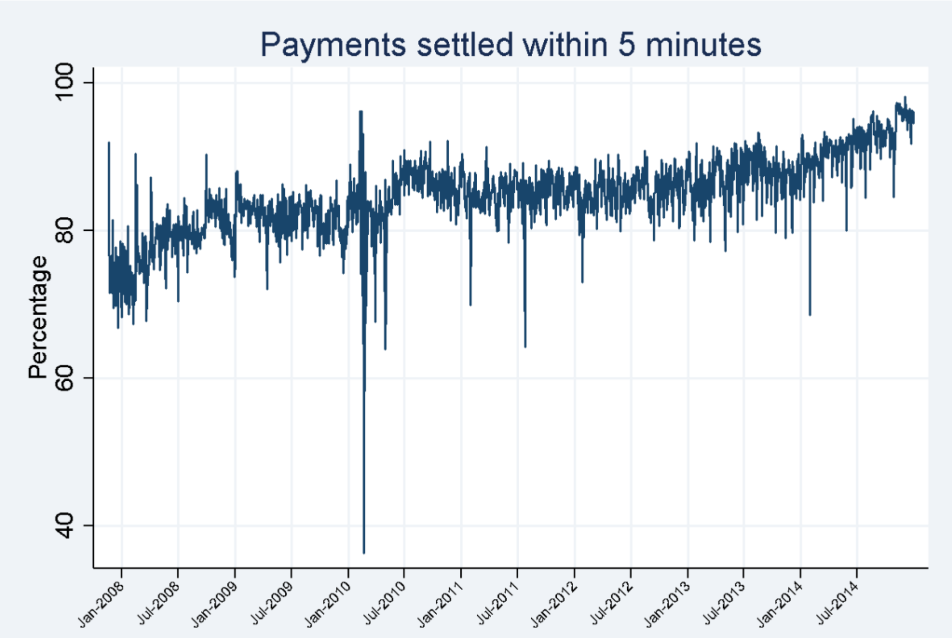
[3] European Central Bank.

The authors are grateful to Matti Hellqvist, Clement Martin, Pierre-Antoine Pineau, and Fabrice Leray for their help in understanding the functioning of the TARGET2 payment system. The authors thank Billiana Alexandrova-Kabadjova and Edoardo Rainone for very insightful discussions as well as participants of seminars at the DNB, the ECB, and Indiana University (Bloomington), and conferences of the Bank of Canada and Systemic Risk Analytics. The help of Monica Reusser is appreciated. The authors of this paper are or were members/alternates of one of the user groups with access to TARGET2 data in accordance with Article 1(2) of Decision ECB/2010/9 of 29 July 2010 on access to and use of certain TARGET2 data. The Banque de France and the PSSC have checked the paper against the rules for guaranteeing the confidentiality of transaction-level data imposed by the PSSC pursuant to Article 1(4) of the above mentioned issue. The views expressed in the paper are solely those of the author(s) and do not necessarily represent the views of the Eurosystem, the Deutsche Bundesbank, the Federal Reserve Bank of Cleveland, the Board of Governors of the Federal Reserve System, the Banque de France or the European Central Bank. Martin Saldias worked on the paper while affiliated at Banco de Portugal.

Non-Technical Summary

In this paper, we focus on the negative externality of free-riding on payment systems using payment-by-payment data from Target2 over the period 2008-2014. We analyze banks' delaying behavior and particularly test two main hypotheses: (i) does a bank react to delays by delaying payments to those counterparties who delay to it, a "strategic reaction" effect? (ii) Does a bank propagate "upstream" delays that are delayed to it, a "pass-through" effect? Our findings suggest that delays do not result from banks' bilateral punishment game but rather from employed intraday liquidity management practices. At the same time, banks tend to delay more when facing incoming delays.

Figure below shows that from 2008 to 2014, on average, only 85% of daily payments were settled within 5 minutes. The simple existence of such delays in a gross settlement system is already puzzling since all payments are to be settled immediately given that banks have enough liquidity to make their payments.



Source: TARGET2 data. Only interbank and customer payments.

Smooth functioning of financial infrastructure is crucial for the stability of a financial system and transmission of monetary policy. Payment systems play a key role in transmission of liquidity between financial agents. Target2 payment system is the Eurosystem's Real Time Gross Settlement (RTGS) system whose yearly turnover reached €470 trillion in 2015, which is equivalent to 30 times Euro area GDP. RTGS systems have advantage over net settlement systems since they settle payments immediately and irrevocably. This allows for reduction in settlement risk but at the same time brings heavier liquidity requirements for participants. As suggested in the literature, when liquidity is pricy, participants may be willing to delay their payments while waiting for incoming payments. Such behavior in certain circumstances can be very disruptive, for example, problems with transferring liquidity between agents may lead to complete freeze of their economic activity. Delays in payment systems have attracted a lot of attention from central banks and specialists in payment systems due to their potential to threaten system's functioning by provoking gridlocks.

We contribute to both theoretical and empirical strands of the payments literature on delays. First, we confirm that banks do delay payments in TARGET2 payments system over the period from 2008 to 2014, and the volume of delayed payments evolves over time. Second, we question the statement made by the theoretical literature that banks delay payments strategically on a payment-by-payment basis as a response to incoming payments being delayed. And in order to do that, we use insights from the literature on econometrics of networks and propose an econometric approach that allows us to distinguish between two types of responses. Namely, a bank reacts by delaying payments only to those counterparties that delay to her, a "strategic reaction" effect, or a bank simply propagates "upstream" delays that are delayed to her, a "pass-through" effect. We treat endogeneity issues by designing a set of relevant instrumental variables.

Our findings suggest a small statistically significant pass-through effect, of the order of a couple of percent of delayed value. However strategic reaction, during the day, is a minor part of delay decisions. In other words, we find that banks do not delay payments strategically to the counterparties that have previously delayed to them on a payment-by-payment basis. If anything, payment delays seem to be rather a part of banks' integral liquidity management practices and probably made at the beginning of the day when banks decide how much liquidity to provide to the system. While decisions made strategically throughout the day have either a minor effect or are absent. It should be noted that all of these results hold for days where there has been no major breakdown in the system.

Retards de paiement : propagation et punition

RÉSUMÉ

Nous utilisons des données uniques de transactions du système de règlement brut en temps réel TARGET2 pour analyser le comportement des banques en ce qui concerne le règlement des créances interbancaires. Une variable cruciale est le temps qui s'écoule entre l'introduction d'un paiement dans le système et son règlement : le délai de paiement. Les retards représentent un moyen par lequel certains participants peuvent profiter gratuitement de la liquidité des autres. Ces retards sont importants car ils peuvent engendrer d'autres retards, ce qui pourrait conduire à une saturation du système. Nous caractérisons les retards dans TARGET2 et examinons si les retards dans les transactions entrantes entrainent des retards dans les transactions sortantes. Nous distinguons le transfert mécanique des retards et la réaction d'une banque à sa contrepartie en retard, et nous proposons un ensemble d'instruments pour résoudre les problèmes d'endogénéité. Nos résultats montrent que les retards se propagent en aval ; cependant, dans la plupart des cas, l'effet est plutôt limité. En ce qui concerne des stratégies sur la base du paiement par paiement, contrairement à ce que prédit la littérature théorique, les données ne montrent que des preuves très faibles. Cette conclusion ouvre la voie à une recherche plus détaillée selon laquelle banques suivent plutôt des routines de gestion des liquidités récurrentes.

Mots-clés : retard de paiement, régresseur endogènes, liquidité, TARGET2.

1 Introduction

Intraday liquidity drives the yield curve at its shortest maturities, and in the payments systems banks respond to their intraday liquidity needs in the most direct way.[1] Information and liquidity needs revealed during the day drive overnight interbank lending, as banks react to payment shocks and reallocate liquidity from cash-rich to cash-poor participants. However, the overnight market is only a part how banks handle a liquidity shortfall since banks can also implicitly borrow from the payments system when they delay the clearing of their outgoing payment until incoming payments clear. In other words, a payment order sent but not secured by liquidity on the bank's account amounts to a free loan and is expressed as a delay. The settlement of a payment happens when the bank either provides additional liquidity to the system or when it receives sufficient liquidity from incoming payments. The liquidity provided by the bank is analogous to a reservoir which the bank tops off when the outgoing payments exceed the incoming ones and the reservoir is too low. Handling the reservoir provides the link between implicit intraday loans (through payment delays) and the overnight market at the slightly longer maturities.

However, this link between the markets for intraday and overnight loans differs from the link between markets of overnight and short-term (e.g., 1 week) loans. A bank deciding whether to borrow from the market balances the interest rate cost of the current overnight loan with the implicit cost of delaying. Such costs include irritated customers, costs due to missed deadlines, or penalties imposed by the participants within the payments system. Further, a payment delay passes down in a mechanical fashion to other participants in the system: non-delaying participants rely on the incoming liquidity to make their own payments and may delay when do not receive it; so these delays can indeed induce other delays and so forth to the extent that the entire system collapses. Such externalities are not included in the banks' calculation of private cost of delaying a payment versus an overnight loan cost. Buckle and Campbell [2003] suggests that non-delaying participants may develop mechanisms that mitigate these externalities. One such mechanism is to delay to those counterparties who delay to them.

In this paper, we analyze delays as a negative externality of free-riding on the system in a reduced-form econometric setting. We establish two facts: first, in normal times, the externality from the propagation of delays is small and significant; second, delays as reactions to those counterparties who delay outgoing payments are extremely small. To establish these facts, we have to surmount two obstacles. First, our data set is every payment processed in the Eurosystem's Real Time Gross Settlement (RTGS) system TARGET2. The system includes different internal mechanisms to improve efficiency and save liquidity, which makes it difficult to measure delays that correspond to the theoretical notion of delays needed to appropriately test our hypotheses. We discuss difficulties of measuring intraday borrowing through delays and define those delays that are most appropriate to intraday liquidity in the RTGS system. We also discuss briefly other mechanisms that could be employed by participants to punish delaying behavior and show that these are essentially not utilized. The second obstacle is that delays as reactions to incoming delays are difficult to distinguish from delays that are simply propagated downstream in a mechanical way but also from

[1] See, for example, Abbassi et al. [2015], who explicitly compute intraday interest rates.

unobserved shocks that affect both participants involved in the transaction. To overcome this issue, we design a set of instruments that make our estimates of reaction-effects consistent under a set of plausible assumptions. These instruments are inferred from the network constructed from a time slice just before a transaction takes place, so that rapid developments in the network prior to the observation are fully incorporated into the instrument.

The negative externality of free-riding on the system is at the core of our interest in studying payment systems since the smooth functioning of financial infrastructures is crucial for the stability of a financial system and the transmission of monetary policy. Payment systems play a key role in the transmission of liquidity between financial agents.[2] Yearly turnover of the TARGET2 payment system reaches €470 trillion, which is equivalent to 30 times the Euro area's GDP. Problems with transferring liquidity between agents may lead to complete freeze of their economic activity. Delays in payment systems have attracted a lot of attention from central banks and payment systems' specialists (see, for example, Bech and Soramaki [2002], Galbiati and Soramaki [2011], Beyeler et al. [2006]) due to their potential to threaten the financial system's functioning by provoking gridlock.[3]

The risk of a gridlock is specific to RTGS, to which many central banks have moved from net settlement systems in order to reduce settlement risks (unwinding of net positions) in the last two decades. The elimination of settlement risk is possible because RTGS systems settle payment irrevocably, and with finality, on an individual gross basis in real time. However, the elimination of this risk comes at a double cost: heavier liquidity requirements for participants and the presence of a central bank to smooth synchronized payment flows; it also creates a potential risk of a gridlock if participants (banks) fail to provide enough liquidity in order to make their payments. RTGS systems are designed in a way to minimize these costs, in particular, different algorithms such as queueing and liquidity-saving mechanisms are put in place in order to improve the efficiency of the system and to economize on the overall liquidity used. And participants are provided with various instruments to manage their liquidity such as beginning-of-the-day balances, credit lines at the central bank available against pledged collateral and bilateral limits. However, the smooth functioning of the system remains dependent on banks' collective behavior, while delaying a payment (waiting for an incoming payment before sending an outgoing payment) remains a beneficial strategy for an individual bank.

The theoretical literature has analyzed banks' delaying behavior and attempted to answer the question which conditions make banks prone to delay their payment. From a game-theory perspective, Bech and Garratt [2003] and Bech [2008] characterize the interaction between intraday liquidity management and payment delays as a coordination game and provide a rationale for the timing of payments and delays. The authors argue that banks engage in a prisoner's dilemma

[2] See Manning et al. [2009] for a comprehensive summary of theory and practice of large-value payment systems (LVPS), including delays. Rochet and Tirole [1996] provide additional insights on net settlement systems and payment systems design.

[3] Bech and Soramaki [2002] define gridlocks as settlement queues where the formation of queues can be attributed to the requirement for payments to be settled individually; in this case, netting of payments may resolve the problem. While gridlocks appear if the formation of queues can be attributed to a lack of liquidity, only the inflow of liquidity into the system can resolve the situation.

game when the central bank's intraday credit policy is to provide liquidity (credit) against pledged collateral. Banks have an incentive to postpone payments since daylight liquidity is costly; however, this is not socially efficient. Buckle and Campbell [2003] show in a theoretical model that delays in an RTGS are likely to occur if banks care about bilateral payment imbalances. The main assumption in this literature is that banks make decisions to delay a payment on a payment-by-payment basis and do it in anticipation of their counterparties behaving the same way.

The empirical literature has documented the existence of delays. Massarenti et al. [2013] provide the first and very thorough characterization of the intraday patterns of payments in TARGET2 between 2008 and 2011 and find that delays respond to timing clustering, which indicates strategic liquidity management behavior on the one hand and contexts in which payment delays might be prone to creating systemic liquidity distress on the other. Bartolini et al. [2010] match brokered trades and Fedwire payment orders and provide a thorough analysis of payment delays. The authors also find that payment delays can be, to some extent, predictable due to their time clustering and therefore they can trigger high-frequency liquidity management decisions to counteract resulting liquidity shortages. They also identify different strategies of market participants, such as the preference of delays for large transactions relative to small trades or, to a lesser extent, delaying settlement when liquid balances are low. Benos et al. [2012] address payment delays in the British RTGS or CHAPS in the aftermath of the Lehman Brothers bankruptcy as an exposure to counterparty default risk in a context of abundant market liquidity.[4] Finally, Heijmans and Heuver [2014] analyze the Dutch part of TARGET2 and suggest that delays in payments to and by a particular bank may be a signal of a bank's stress.

We contribute to both theoretical and empirical strands of the payments literature on delays. First, we confirm that banks do delay payments in TARGET2 over the period 2008 to 2014, and the volume of delayed payments evolves over time. The data section provides a thorough analysis of delays with a particular focus on the appropriateness of the data definitions for the subsequent econometric exercise. Second, we investigate the statement made in the theoretical literature that banks delay payments strategically on a payment-by-payment basis as a response to incoming payments being delayed. And in order to do that, we propose an econometric approach that allows us to distinguish between two types of responses. Namely, *(i)* a bank reacts by delaying payments only to those counterparties that delay to it, a "strategic reaction" effect, or *(ii)* a bank simply propagates "upstream" delays that are delayed to it, a "pass-through" effect. We design a set of instrumental variables that allow us to handle endogeneity issues. After designing a relevant set of instruments, we conclude that there is a small pass-through effect that is statistically significant, usually of the order of a couple of percent of delayed value. However strategic reaction, during the day, is a minor part of delay decisions. The coefficients of delay responses to purely mechanical parts of the system, such as whether the system is flooded with many payments to process, are much more economically significant. In other words, we find that banks do not delay payments strategically to the counterparties that have previously delayed to them on a payment-by-payment basis. If anything, payment delays seem to be rather a part of banks' integral liquidity management practices and probably made at the beginning of the day when banks decide how much liquidity to

[4]The authors also provide an alternative definition of delays to the one studied in this paper.

provide to the system. While decisions made strategically throughout the day have either a minor effect or are absent. It should be noted that all of these results hold for days where there has been no major breakdown in the system.

The remainder of the paper is organized as follows. In Section 2, we provide an overview of TARGET2 data and document intraday patterns of delay. Section 3 develops the econometric framework we use to analyze the propensity of incoming delays to propagate in payments downstream. Section 4 discusses econometric results and provides evidence on the use of bilateral limits by banks. Section 5 provides details on the robustness checks. Finally, we conclude and discuss future research in Section 6.

2 Data

In this paper, we study delays in the large-value gross settlement payments system TARGET2, as they have highly relevant implications for systemic liquidity risks and the early identification of potential changes in market participants' behavior. Unlike other studies, we measure a payment delay precisely, namely as the difference between a payment's introduction into the system and its settlement, calling any difference greater than 5 minutes a delay.[5,6] Figure 1 shows that from 2008 to 2014, on average, only 85% of daily payments were settled within 5 minutes. The simple existence of such delays in a gross settlement system is puzzling since all payments are to be settled immediately given that banks have enough liquidity to make their payments.[7] Our study aims to provide a deeper analysis of the delays themselves and the reasons why they happen.

[5] In our paper, payment delay has a technical meaning, the time needed to settle a payment after its introduction to the system. Benos et al. [2012] and Bech et al. [2008] look at a payment delay from a different angle, the time a bank holds incoming liquidity before sending it out. Since this information is not readily available in the data, they estimate it using Markov chain theory.

[6] A 5 minute interval is considered as processing time Massarenti et al. [2013]

[7] Massarenti et al. [2013] were the first to document this pattern for TARGET2, and we refer the reader to their paper for a thorough analysis of the payments data over time (2008-2011) as well as intraday patterns. However, they did not try to explain why these delays happen.

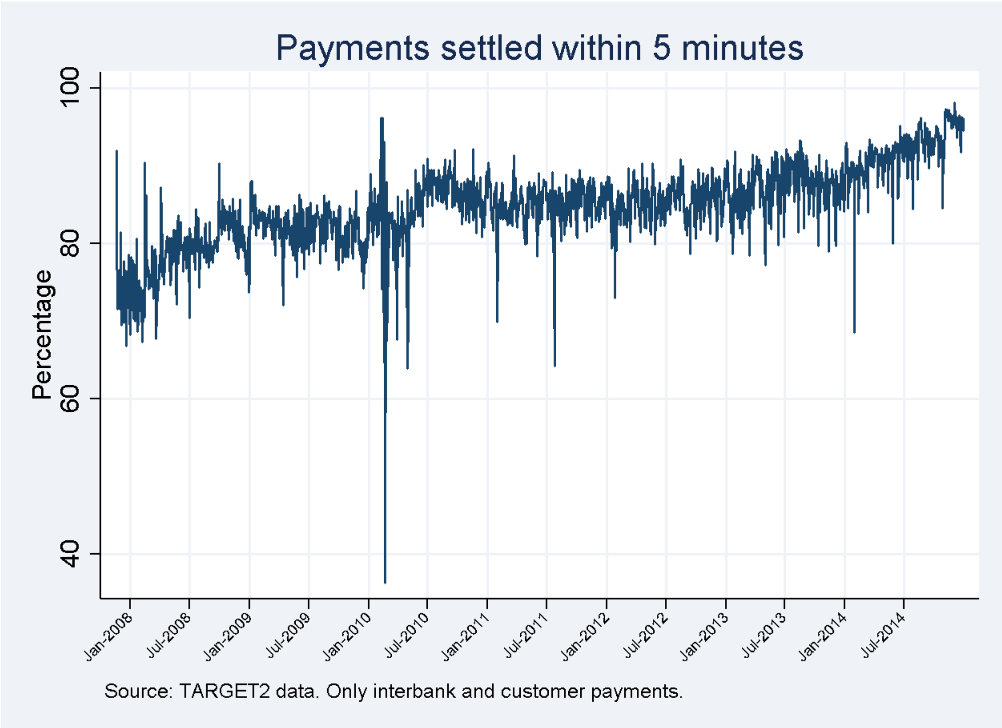


Figure 1: Daily fraction of interbank and customer payments settled within 5 minutes

2.1 Data Definitions

The Eurosystem's real-time gross settlement (RTGS) system TARGET2 started operations on November 19, 2007. All the euro area countries plus Bulgaria, Denmark, and Romania gradually joined the TARGET2 system, and by 2013 it comprised 1700 credit institutions.

Our analysis covers June 2008 to December 2014.[8] All the payments processed in TARGET2 can be divided into four categories: payments between commercial institutions; payments to/between central banks; payments related to the settlement of ancillary systems; and liquidity transfers between payment systems, or between different accounts of the same participants. In Panel A of Table 1, we document daily values and volumes of payments in each of these categories averaged over the period 2008-2014. Payments between commercial institutions that include both interbank and customer payments constitute the majority of the payment volume (3/4 of the amount of daily payments and 1/4 of the value), whereas payments related to the settlement of ancillary systems and liquidity transfers make up most of the daily settlement value (2/3 of the value and 15% of the volume). Payments to/between central banks represent about 9% both in volume and in value.

In an RTGS system, all payments are supposed to be settled immediately given that a participant provides enough liquidity. We are interested in delayed payments that are a part of agents' strategic behavior. In particular, we define a payment as delayed if its execution time, i.e., the difference between the point in time at which it is introduced to the system and the time it is actually settled, exceeds 5 minutes. This definition allows for routine processing time and allows us to focus on those

[8] We exclude from our the analysis the period from November 2007 till May 2008 as a "burn-in" period of the payment system.

	Panel A.		Panel B.	
Category	**Volume**	**Value**	**Volume delayed**	**Value delayed**
Customer payments	57,44%	4,27%	13,41%	8,02%
Interbank payments	18,71%	18,39%	21,19%	9,65%
Central bank payments	9,16%	9,01%	6,79%	6,04%
Ancillary systems	10,61%	20,04%	4,53%	9,19%
Liquidity transfers	4,06%	48,08%	5,33%	1,87%

Table 1: Panel A: Payments by category, average volume and value of payments in the TARGET2 payment system settled during 06/2008-12/2014. Panel B: average value and volume delayed within the category of payments.

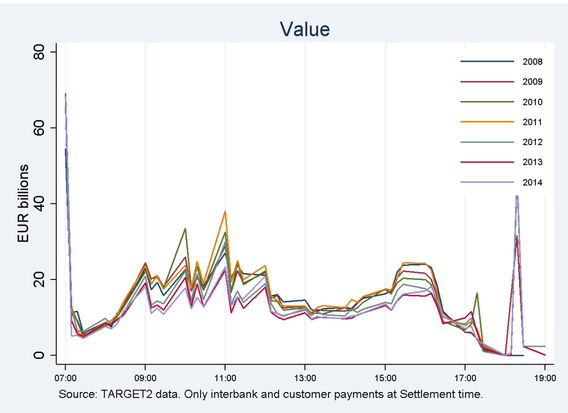
delays that may be the cause of further delays. Introduction to the system is defined by TARGET2 rules as the moment the participant sends a payment message. However, participants are permitted to postpone the introduction time of a payment by specifying an "earliest debit time". For example, they can send a payment message at 7 am to indicate a payment should be introduced to the system at 10 am. In addition, payment messages can be sent to the system outside the settlement hours, i.e., 7:00-18:00 (so-called "warehouse payments"); in this case, a payment will be introduced at 7 am next business day. Taking these details into account, we define the introduction time as the maximum between the time the payment is sent to the system and earliest debit time if the introduction and the settlement of the payment occur on the same business day; or 7 am in a case of warehouse payments.

Delays may happen for several reasons: banks' behavior (reaction; bilateral or multilateral limits reached), liquidity shortage (pass-through effect), or operational issues (e.g., technical slowdowns in systems' performance). We are interested only in the delays that happen due to the first two reasons with which only commercial institutions ("banks" from now on) should be concerned. Indeed, central banks are liquidity providers to the market and therefore have no incentive to delay payments. Ancillary systems mechanically settle payments and have no strategic behavior. Liquidity transfers involve no change of ownership of the money being transferred, and thus such transactions should not be susceptible to being delayed. Panel B of Table 1 lists average daily values and volumes of delayed payments in each of these categories and partly confirms our intuition. Most of the delays happen for payments between commercial banks: on average, a third of the payments made by commercial banks, or 24% of all the transactions, are delayed. We do see though that payments from the last three categories can be delayed as well; however, the number of delayed payments in these three categories together makes up only about 4% of total payments volume, i.e., 16% is delayed out of the 23% of transactions made. One possible explanation for these delays is that they are made during rush hours like 7 am. At this time, payment settlement takes some time due to too many payments being processed at the same time. Liquidity-saving mechanisms used in the system to optimize the use of liquidity could be another reason; however, more analysis should be done for a better understanding. For the rest of our study, we focus only on the interbank and customer payments carried by banks. The proportion of delayed payments in these categories is much more significant, though they may be subject to the same issues. We address both of them in our econometric exercise.

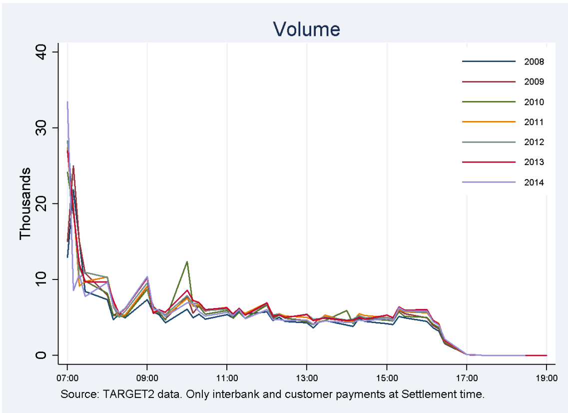
2.2 Intraday patterns

Analyzing the behavior of intraday payments patterns is essential to understanding the underlying liquidity needs of the participants in the payments system. Patterns demonstrate fairly strong consistency from one year to another. Figures 2a and 2b show yearly averaged values and volumes that are settled during each fifteen-minute slot from 7:00 to 18:00 (which is the cut-off for interbank transactions). Such persistence is important for operational and oversight purposes because it allows the detection of any deviation from average behavior.

Both graphs of value and volume exhibit a similar daily rhythm: the biggest spike is at 07:00, the beginning of the trading day, and then there is a significant decrease; several upticks occur in the morning hours, at 09:00 and 10:00 and finally around 16:00. The value curve has some additional spikes at 11:00 and 12:00 and also at the end of the day. These spikes could be due to the liquidity settled by ancillary systems such as securities settlement systems, central securities depository, or central clearing counterparties. The hump occurring around 16:00 corresponds to the the settlement of net positions of EURO1 and other ancillary systems. At the opening of the system, spikes in value and volume are equally important: payments are numerous and rather small. Later during the day, fewer payments are settled but are of higher value; particularly at the end of the trading day, some payments settled are of extremely high value. For more discussion on the intraday behavior of TARGET2 payments, we refer to the paper by Massarenti et al. [2013].



(a) Intraday patterns of yearly average value of payments. Value of payments are computed for 15-minute intervals

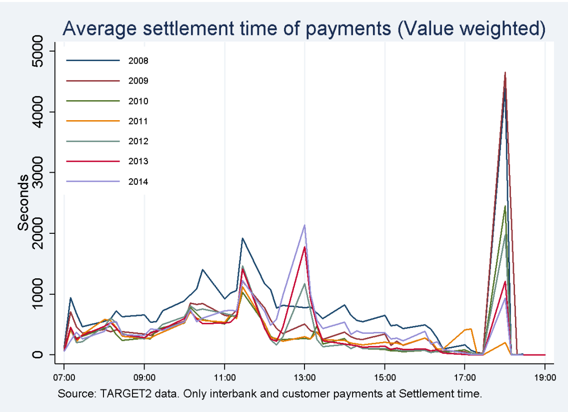


(b) Intraday patterns of yearly average volume of payments. Number of payments are computed for 15-minute intervals

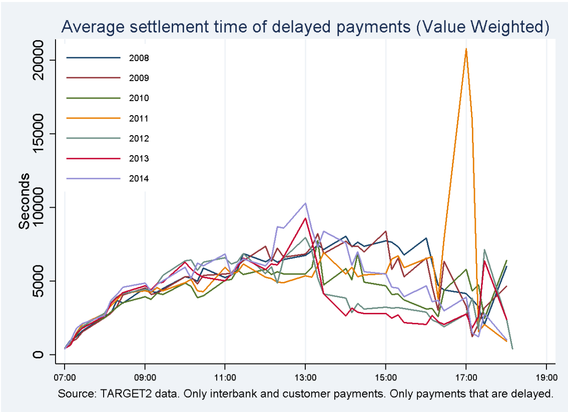
Figure 2: Intraday patterns of payments

Figure 3a depicts an average settlement time normalized by total number of payments and weighted by payment value. We see again similar patterns of the settlement time during the day over the years. Three main spikes are observed during a typical day: at 7 am; around noon; and at the end of the trading session. The charts for all years exhibit similar patterns with two exceptions: first, the magnitude of the whole curve is somewhat distinct in 2008; second, in 2008-2011, settlement delays built up till 11:30 and then slowly faded away, and since 2012, another spike appeared around 13:00 and faded away at a faster pace. Finally, an important spike appears right before closing of operations that is due to the few payments introduced and to their large value. We plot charts of value-weighted settlement time because it allows us to take into account the fact that a 1-hour delay of a payment of 100€ and one of 1€ billion may have different consequences for the system. At the same time, one should keep in mind that very large-value payments could distort the average delay upwards. The daily maximum time needed to settle a payment does not exceed 2000 seconds on average (a bit more than 30 minutes) except for the spikes at the end of the day, particularly for years 2008 and 2009 when this time reached 4500 seconds (1,5 hour). We note that the magnitude of the series is decreasing through time, suggesting higher efficiency of the settlement process. This could be due to two main reasons: better liquidity management by banks (learning curve) and liquidity injections by the central banks.

Since delayed payments represent a rather small fraction of all payments, normalization of settlement time by the total number of payments can significantly average out this information. Therefore, Figure 3b depicts an average payment delay normalized by the number of delayed payments and weighted by payment value. We see that the delay time is much higher, being on average 5000 seconds (about 1,5 hour), and the general pattern resembles a reverse U-shape curve, with the length of delays gradually increasing during the first half of the day and reaching its maximum around 1 pm and then slowly decreasing. This U-shape curve is simply an accentuated form of the build-up of the delays over the first half of the day observed in the previous graph. Interestingly, there is much less consistency over the years, which, from our point of view, underlines the nonmechanical nature of the existence of delays.



(a) Intraday patterns of settlement length of payments weighted by the payments value. Normalized by the total number of payments. Computed for 15-minute intervals

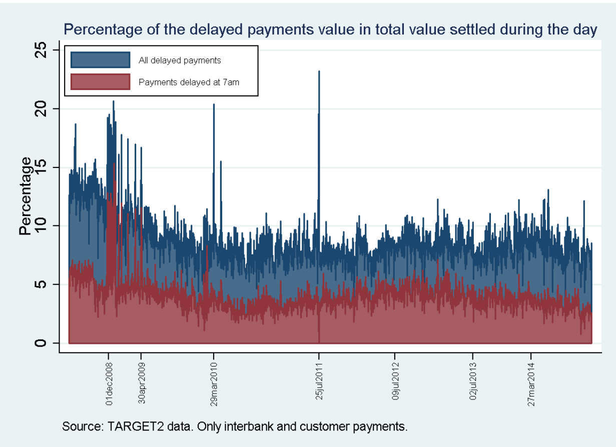


(b) Intraday patterns of settlement delay of payments weighted by the payments value. Normalized by the number of *delayed* payments. Computed for 15-minute intervals

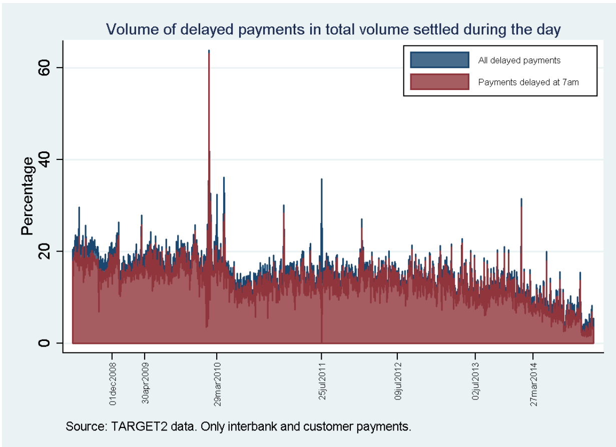
Figure 3: Intraday patterns of payment delays

2.3 Morning queue effect

The TARGET 2 Annual Report (ECB [2013]) mentions the delays in the first hour of trading days, the so called "morning queue effect", explaining this phenomenon by the fact that too many payments have to be settled in the first hour and therefore the processing of these payments takes more time. Indeed, in Figure 2b, we can see that the number of payments sent to the system is about two to three times higher at 7 am than at any other time during the day. However, payments are smaller and processed relatively faster with respect to payments delayed during the day (see Figure 3b). To develop further the importance of the morning queue effect, figures 4a and 4b display average daily values and volumes of payments delayed both at 7am and during the whole day. We can see that on average early-morning delays constitute the majority of the daily delays in terms of the volumes whilst intraday delays make up about half of the total delayed values. In the report, the first hour of operations is excluded from the statistics on TARGET 2 processing times. However, Massarenti et al. [2013] base their descriptive analysis on all the delays observed during the day. In our study, we are interested only in delays that result from banks' strategic behavior; therefore, we consider it important to take this specificity into consideration and perform our econometric analysis on payments excluding those that take place at 7 am.



(a) Percentage of the daily value delayed at 7am and during the rest of the day



(b) Percentage of the daily volume delayed at 7am and during the rest of the day

Figure 4: Daily value and volume of 7am and during-the-day payments

2.4 Bilateral limits

Bilateral limits are a tool provided to banks to control their outgoing payments. Banks can put a cap on the total value of transactions with each counterparty or with respect to all counterparties at the same time. This cap can be set at any time during the day and can also be set beforehand to take effect at a specified time. Given this, one might expect bilateral limits to be used as a disciplining tool. If a counterparty is perceived as a free rider, then a bank might enforce a bilateral limit against it so that even automatic payments do not go out to the offending payee. However, strategic or not, these limits are never used by the vast majority of banks. Indeed, of the over 1700 banks using the system during the 7-year period of our data, only 24 banks ever used the limits. Even among the

24 banks that used the bilateral limits, there is some evidence that, for many of these, they used the limits largely as a test, and then did not use them again. For example, the limits were set on a one-time basis at an extremely high value compared to the total payments to that counterparty, and then after several days they were removed and never used again.[9] This suggests a finding that we formalize in our more structural empirical work: on a payment-by-payment basis, the agents in the system do not behave strategically. If they strategically react to a specific counterparty who has a shortage of liquidity or exhibits free riding through payment delays, it is likely that they enact other strategies. In this case, individual agents do not take on the role of system discipline through the use of bilateral limits.

3 Econometric model of delay drivers

In this section, we focus on distinguishing between two externalities due to delays in payments data using an econometric approach. The first one is a delayed payment made to a free rider in response to its delayed payment. This response could be a punishment mechanism or a response to information given by this particular delayer. Second, upstream delays can provide an unexpected lack of liquidity to participants who, in the absence of the delays, would have enough liquidity to clear their payments in time. This reason is of special interest to economists in that a measure of this effect gives an idea of an external cost to the system from the free-riding borrower who does not carry enough liquidity. Our tests will focus on these two aspects of the fallout from the intraday borrowing that implicitly occurs when a participant decides to "let a payment ride" rather than top off its expensive liquidity.

We look at the behavior on a variety of days with the idea that in a collateralized credit regime banks have incentives to postpone payments when daylight liquidity is costly. Banks are, therefore, expected to delay payments as a reaction to incoming delays, and even more so, during periods of scarce liquidity. The reaction to both incoming payment delays and to the information implicit in a delayed payment (whether the payer is a "free rider" or other information) should be more pronounced on days when liquidity is perceived as scarce.[10]

We test these hypotheses using transaction-by-transaction payments data from TARGET2. In the following, our dependent variable is a delay in a transaction from bank i, the "sender", to bank j, the "receiver". In a general form, the problem can be formulated as follows: when the sender bank i receives a payment or a series of payments that is delayed, to what extent does this cause the sender to delay its payments to other banks? As mentioned above, we distinguish two types of hypotheses: *(i)* a sender's strategic reaction to delayed payments, that is, if a bank reacts to incoming delays by delaying more to those peers that delay to it; and *(ii)* a simple pass-through of payment delays. In the case of a pass-through, a bank's response to delayed payments could be

[9]There is one significant outlier bank which puts, on average, 1400 limits against 700 counterparties per day. However, most of the 24 banks set fewer than 100 limits and with respect to a dozen counterparties over the entire period. Diehl [2013] also reports few banks using bilateral limits.

[10]For example, Bech and Garratt [2003] and Bech [2008] posit simple models where payments delays are time-varying.

mechanical in that no strategic behavior is involved. This question assumes a certain importance if delays may propagate causing the entire connected network to be in delay. Our estimates should uncover the extent to which the delay propagation can intensify.

More formally, in our empirical analysis, our level of observation is a single payment made by bank i to bank j at time t_0. We study the information content that is in the delays of payments to the paying bank i during a time interval $t \in [t_0 - s, t_0)$ just before the observation; and we denote F_m^t as the information available to participant m at time t. We start by analyzing hypothesis *(i)*, a sender's strategic reaction to delayed payments. We are capturing the information contained in delays by the receiver of the current payment j in a variable which we denote with a subscript R for *"Reaction"* because this variable measures the delays that the payer directly reacts to in response to the past history of payments from this particular receiver, bank j. (We use the term reaction to include a possible punishment for free-riding.) To capture the response of bank i to delays propagated from "upstream" payments, we use a set of variables measuring payments initiated to bank i during the preceding period $[t_0 - s, t_0)$. In other words, if $j \in J$, $j \neq i$, and where J is the set of all payers who paid to i during the period, then $LogInValue_{Ji}^{t_0-s,t_0}$ represents the logarithm of the total value of payments to i during the period. Similarly, $LogOutValue_{iK}^{t_0-s,t_0}$ represents the logarithm of the total value of outgoing payments of i during the period. A typical regression model is

$$\begin{aligned} Prob(Delay_{ij}) = \alpha(ProportionDelayed_{Rji}^{t_0-s,t_0}) + \beta LogValueDelayed_{Ji}^{t_0-s,t_0} + \\ \gamma LogInValue_{Ji}^{t_0-s,t_0} + \delta LogOutValue_{iK}^{t_0-s,t_0} + Controls + e_t \end{aligned} \quad (3.1)$$

In this case, we measure the delay response of bank i in paying bank j at time t_0 as a probability that the delay occurs. It is a linear function of the percentage of the value that bank j delayed directly to bank i during the time period $[t_0 - s, t_0)$ just before bank i's payment is made. It is also a linear function of the log of total delayed payments and the total incoming and outgoing value of payments to and from bank i during the same period. As described above, each payment that is not settled within a 5-minute interval after its introduction to the system is considered to be delayed.

To test hypothesis *(ii)*, we use the following regression, where the dependent variable is the amount delayed (in logs) by bank i to bank j. It takes a value 0 if the payment was not delayed and the value of the payment if it was delayed. In this case, we are interested in measuring if bank i delays bigger amount to bank j at time t_0 if it faces a larger volume of incoming delayed payments during the time period $[t_0 - s, t_0)$.

$$\begin{aligned} LogValueDelayed_{ij} = \lambda\, LogValueDelayed_{Ji}^{t_0-s,t_0} + \mu\, LogInValue_{Ji}^{t_0-s,t_0} + \\ \nu\, LogOutValue_{iK}^{t_0-s,t_0} + Controls + e_t \end{aligned} \quad (3.2)$$

Controls in both types of regressions include a log number of payments and the proportion of high-priority payments during the last 5 minutes. Since a payment system follows certain settlement algorithms, and payments may be cleared with a delay due to functioning reasons, namely numerous payments sent at the same time or multiple high-priority payments in a queue, we want to control for this. Therefore, we introduce our two control variables: the log number of payments and the proportion of high-priority payments during the last 5 minutes. The rationale behind the latter control is the following. All the payments are split in three categories by their priority, and this is not a bank decision. Payments with higher priority are settled first, while others may wait in a queue. Since all the interbank payments have the lowest priority, they may be delayed when many high-priority payments are about to be settled at the same time.

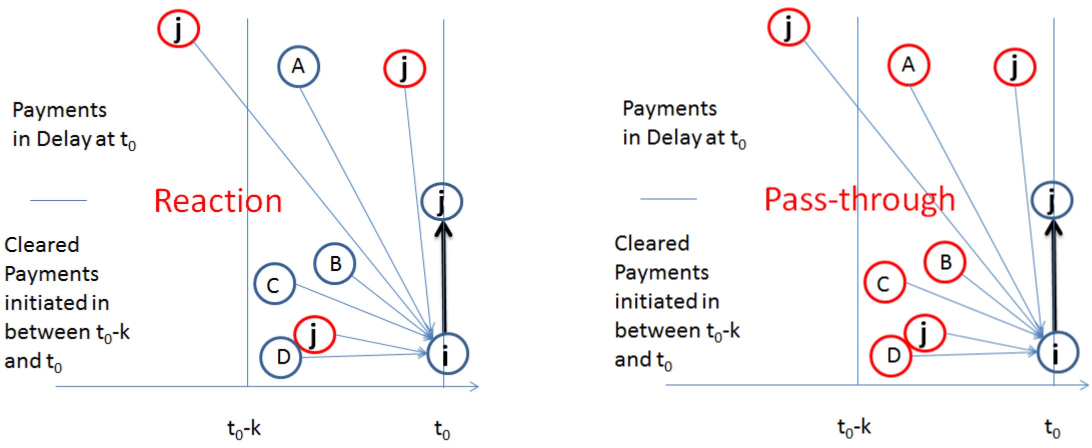
Our estimation considers each transaction as occurring in a sequence of networks. In other words, we choose a time increment, s, and then consider those transactions that occur in the time interval $[t - s, t)$ when looking at those variables and instruments that influence the transaction through the payments network. Because of this, transactions take place in the context of a sequence of rolling networks where the relevant network is the one established by transactions that precede the payment most closely in time. The choice of the size of the rolling time slice, s, depends on a tradeoff. A narrower time slice makes sure that the network implied by the slice is more recent and thus more relevant, which is important when the network could be changing rapidly during the day. However, a too narrow time slice carries disadvantages. The network might not be as complete as it should be because relevant connections are omitted. Further, both variables and instruments that are based on averages of transaction variables will be noisier, so that the instruments will have less power.[11]

The rolling time slice has advantages over viewing the day as split into one-hour intervals. By splitting the day into fixed intervals, consideration of those transactions that precede a payment made near the end of an hour would include transactions that occurred almost two hours prior, but would exclude those transactions that occur less than one hour prior to it. This could mean that essential information is lost in the noise in a period of rapidly changing network patterns. The rolling time slice can also emphasize the time series nature of our observations. Because of this we are careful to report standard errors that are adjusted for the time series properties of the data. The time series nature of the transaction data may also help with the problem, observed by Fox [2008], of the many biased and inconsistent estimates that can result when one draws observations from a single network. By viewing the transaction as occurring within a sequence of time series networks (albeit ones that are serially connected), some of this may be mitigated.

Under model 3.1, the coefficient α captures the influence of delays by a bank's peers on bank i's decision to delay its payments to them (see Figure 5a for a graphical representation). The hypothesis suggested by the theoretical literature is that banks do take strategic actions to delay their payments, i.e., $\alpha > 0$. Under model 3.2, the coefficient of interest is λ, which measures a pass-through of payment delays (Figure 5b). This is nonstrategic component, and we expect it to be positive, meaning that banks do delay more when payments to them are delayed. However, if

[11] We also considered intervals where the interval was defined by the number of transactions, rather than number of seconds. This had the advantage of defining the relevance of the recent network in terms of the information available to the network, if information available is defined in a way that it is proportional to the number of transactions. However, this clashed with our variable definitions, which were measured in terms of transactions per unit of time. If network instruments are measured per transaction-based-interval, then keeping the different units consistent could be confusing, so we report our results for where the time slice is defined in time.

this coefficient is greater than one, then the value of delayed payments increases the amount of further payments being delayed by more than the initial amount, and the whole network, under certain conditions, may get stuck in delays. Total incoming and outgoing values are supposed to control for banks' activity, and we expect their corresponding coefficients to be positive and negative, respectively, since incoming payments should increase the reservoir and facilitate outgoing payments, whereas outgoing payments should have the opposite effect.



(a) Reaction **(b)** Pass-through

Figure 5: (a) Reaction. How much more likely is a payment from bank i to bank j to be delayed given the history of payment delays going from bank j to bank i? (b) Pass-through. Will bank i delay payment to bank j at time t given the history of all incoming and delayed payments? In other words, will an increase in a fraction of incoming delayed payments over all incoming payments lead to a higher probability of bank i delaying its own payment?

A problem we confront in assigning economic interpretations to the parameters α and λ is that the economic value of these parameters depends on the specificity of the information conveyed implicitly from j and J to i, through the delayed payments. For example, α is valuable to research in that it shows whether payer i is reacting to the delay either through punishing the free-riding j who conveys that he has put too little liquidity in his account for payments clearing, or he conveys that he is a bad risk because he is short on liquidity. Similarly, λ has value in that it shows how i reacts to the information that upstream providers are short of liquidity, either passively by allowing the delay to continue downstream, possibly causing more delays down the way, or actively, either by adding more liquidity to the system that needs it, or by not adding liquidity that might normally be added as i adopts a "wait and see" posture to the liquidity shortage. In each of these cases, λ adds to knowledge of the propagation of the delays throughout the network. To clarify, if F_{mt} is the information available to participant m at time t, then 3.1 can be rewritten as conditional on this information:

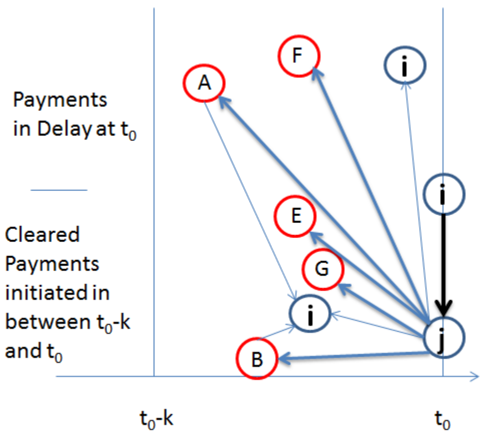
$$Prob(Delay_{ij}|F_{it}) = \alpha(ProportionDelayed_{Rji}^{t_0-s,t_0}|F_{jt}) + \beta LogDelayValue_{Ji}^{t_0-s,t_0}|F_{Jt}+ \\ \gamma LogInValue_{Ji}^{t_0-s,t_0} + \delta LogOutValue_{iK}^{t_0-s,t_0} + Controls + e_t \quad (3.3)$$

Under this formulation, it is likely that the error term will incorporate information that is locally available to i, j and J, inducing a bias. The weakness of a simple regression lies in the endogeneity problem that leads to biased and inconsistent estimates of the parameters of the model: in fact, peers' delaying behavior affects the decision of a specific bank to delay its payments or not, but this bank's delayed payments may also in turn affect the choices made by its peers. In other words, information that is held by i at some time earlier than t_0 is passed to j and J, which is then passed to i. This is a known reflection problem formulated by Manski [1993], which makes empirical identification of effects embodied in α and λ challenging. The problem arises from the fact that the ratio of delayed payments, $ProportionDelayed_{Rji}^{t_0-s,t_0}|F_{jt}$, is itself an endogenous explanatory variable since it is determined simultaneously with the outcome variable.

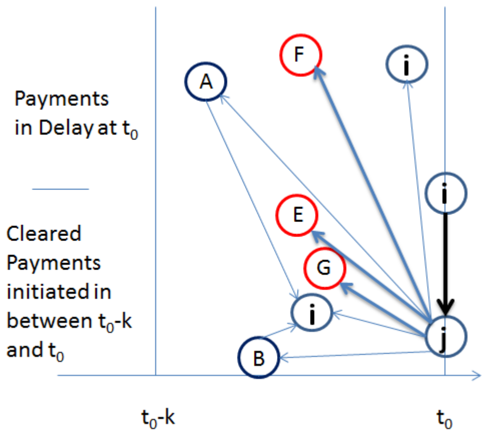
In the case of interactions that are structured in a network, Bramoullé et al. [2009] propose a static solution to this problem in a linear-in-means setting. More particularly, in a network, two connected banks may have different communities of peers, and this intransitivity in network connections can be used as an exclusion restriction to identify the peer effects of interest. This approach suits our story since heterogeneity in bank's decisions with respect to its peer group allows us to use payments delayed *to* the bank's peers group *but* not to the bank in question as a relevant instrument to capture behavior of any given bank with respect to its delaying counterparty.

To illustrate, consider Figure 6a. While making a decision to delay to bank j, bank i knows how much bank j is delaying to it. Bank j's peer group includes banks to which bank j delays except i. Therefore, we can consider the proportion of payments delayed by bank j to all its peers *but* i over the total payments outflow to the same group as a valid instrument for bank j's delays to i, satisfying both the relevance and exclusion conditions. Indeed, this instrument is *relevant* for bank j's decision to delay to bank i, and at the same time, it affects bank i's decision only through bank j (exclusion condition).

To deepen our analysis, we use another set of instruments that allows us to address the errors mentioned in Blume et al. [2010], where a pair of nodes is more likely to delay together. Either because of proximity, similar business models, or exposure to risk from the same funding sources, this pair of nodes share a common unobserved local factor that drives both of them to delay. We address this problem in a similar way, where we use the nodes that are not immediately connected to node i as instruments for the delays by node j. The way the instruments are computed is the same as in the previous case; however, peer groups are defined differently. Particularly, bank j'peer group consists of those nodes that j is connected *to* but *only* through intermediaries that themselves are not linked to bank i. In other words, to construct the second set of instruments, we exclude both direct payments going from j to i and also all the payments going from j to i through first-layer intermediaries (Figure 6b).



(a) Instrument set 1.



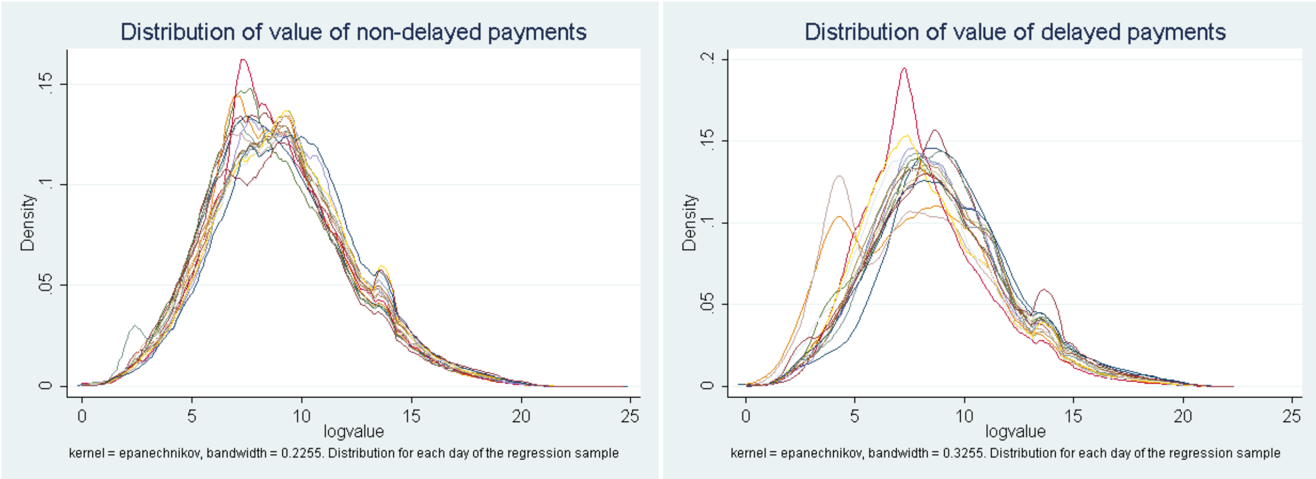
(b) Instrument set 2.

Figure 6: (a) Instrument Set 1. (b) Instrument Set 2. We are interested in analyzing the response of bank i to an incoming delayed payment from bank j. The issue of endogeneity is addressed through the use of instrumental variables. Instrument Set 1 is defined as payments delayed by bank j *to* its peers *but* not bank i. Instrument Set 2 takes into account the possibility that a pair of nodes delay together and defines j's peer group as nodes that j is connected *to* but which are *not* themselves linked to bank i.

4 Results

We perform our analysis on twenty-one days. We chose those days so that ten were "normal" (average stock returns, volatility, and volumes in European markets); five were when markets were "stressed"; and 6 had either particularly high volumes of payments within the day or very few delays. Separately, we analyse 4 more days with a very high number of delays. The reason that we do not run regressions on every day from 2008 to 2014 is that the transaction-level data are computationally intensive. On average, we have more than 300000 transactions per day, and instruments have to be computed for each transaction. As a robustness check, we run the same regressions for all the twenty-five days together using additional dummies per day. The results are similar in sign and size. The sample of the days can be seen in the figures with plotted regression coefficients (e.g., Figure 8a).

We start by testing whether banks have any strategy of delaying payments based on their transaction value. Figures 7a and 7b depict distributions of log values of nondelayed and delayed payments for each day of the sample. We can see that there is again a lot of persistence from one day to another. Moreover, distributions of values of delayed and nondelayed transactions are very similar, with a bit bigger variability of the delayed payments. This suggests that banks have no strategy of delaying payments based on the transaction value.



(a) Distribution of value (log) of non-delayed payments for each day of the sample

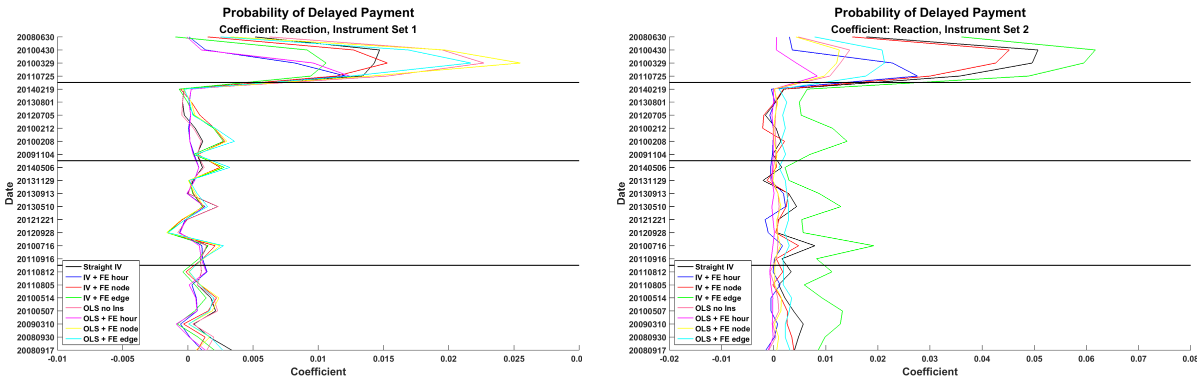
(b) Distribution of value (log) of delayed payments for each day of the sample

Figure 7: Distribution of payments values

Figures 8 and 9 show our results for each of the twenty-five days. These figures the coefficients for each of the estimations for each of the days in our chosen sample. The days proceed down the vertical axis and are clustered by the type: days corresponding to volatile financial markets are in the lower part of the graph; they are followed by calm days and days with few delays and many payments; lastly, 4 days with an extreme number of delays are at the top of the graph. The magnitudes of the coefficients are detailed on the horizontal axes of each graph. Each coefficient is displayed for a list of several instrument combinations. The black line represents ordinary least squares, the dark blue lists coefficients from the instrumental variables as discussed in the econometrics section above without the fixed effects, and three fixed effects regressions are also included: hour, nodal, and edge-fixed effects. The reaction estimate is graphed for two sets of instruments, one that is one link removed from the immediate upstream link and one that is two links removed, as discussed in the preceding section.

Analyzing the reaction coefficients in Figure 8, several things are worth noting. First, for most of the days, coefficients are significant and mostly positive but very small, not exceeding 0.3%. Second, each day is somewhat special, and coefficients vary over time; however, variation magnitude depends on the specification. Third, all the estimates are very similar except for the upper part of the graph featuring days with a lot of delays. We believe that the specification with fixed effects by hour is the best one since it allows controlling for the within-the-day-clustering nature of payments observed previously. These coefficients are mostly the smallest and also the least volatile over time. Nodal and edge effects tend to amplify the reaction coefficient probably capturing some within-node/edge specificities. Interestingly, instrumental variables do not improve significantly over the OLS when using the same fixed effects. Fourth, on the days with many delays, the coefficients may reach 1.2% - 2.7%, and the coefficients are again the smallest when controlling for settlement hour. Finally, it is interesting to note that the biggest dispersion in coefficients is on both market and payment system stress days, suggesting that different factors may be at play at the same time.

All in all, our findings suggest that banks do not delay payments to the counterparties that delay to them, since the probability of a such a delay is on average 0.1%. The results are very similar for the second set of instrumental variables but with some more dispersion across specifications.



(a) Instrument Set 1 **(b)** Instrument Set 2

Figure 8: Reaction. Does a bank react to incoming delays by delaying to those counterparties that delay to it? (a) Instrument Set 1: Excluding payments coming from the bank in question. (b) Instrument Set 2: Excluding payments at the second layer passing through the bank in question. Horizontal lines separate types of days: the lower chart corresponds to market stress days; it is followed by market calm days; then days with extremely many payments or few delays; and on the top, days with extremely many delays.

Figure 9 shows the pass-through coefficients. The interpretation of the graph and conclusions are very similar to Figure 8. First, coefficients on most of the days are both positive and significant, but also higher, on the of order 2%-5%. Second, among all specifications, estimates with hour fixed effects are the smallest. However, there is more dispersion of estimates across econometric specifications and over time. Third, the coefficients become very large on the days with multiple delays. We will discuss this phenomenon in a while.

We conclude that banks do tend to delay more if they have more delayed incoming payments. However, since the coefficients are rather small on most of the days, it indicates that the propagation of delays is not long lasting.

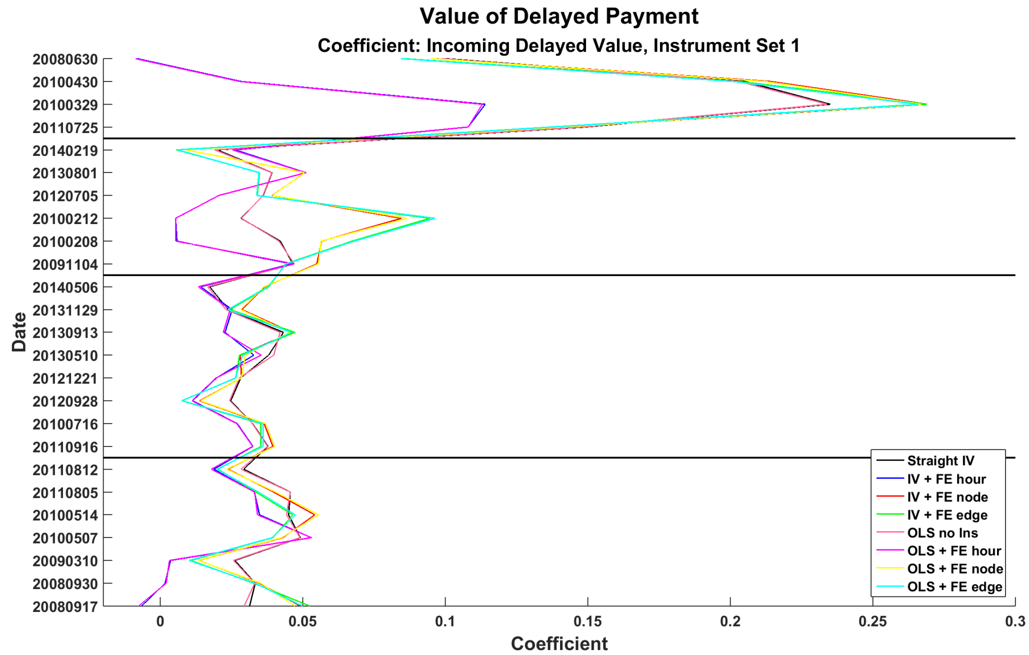
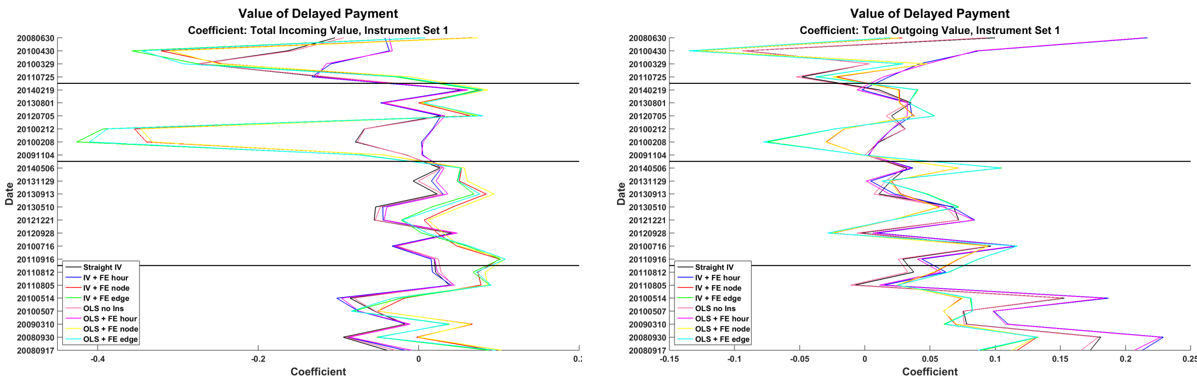


Figure 9: Incoming delayed value. Horizontal lines separate types of days: the lower chart corresponds to market stress days; it is followed by market calm days; then days with extremely many payments or few delays; and on the top, days with extremely many delays.

In Figure 10, we plot how total incoming and outgoing value impacts the likelihood of an outgoing payment being delayed. Coefficients are of the expected sign, mainly negative and significant for incoming value and mostly positive and significant for outgoing value. This is consistent with our story that it is the ratio of delayed payments that matters. The coefficients though vary a lot across specifications and over time.



(a) Total incoming value **(b)** Total outgoing value

Figure 10: Total incoming and total outgoing value. Horizontal lines separate types of days: the lower chart corresponds to market stress days; it is followed by market calm days; then days with extremely many payments or few delays; and on the top, days with extremely many delays.

To understand the size of the pass-through and reaction coefficients on certain days, we plot the fraction of payments delayed during 15-minute slots for these three days in the middle of our sample:

the three crisis days and the one normal day (see Figure 11). The picture is self-evident: on each of the crisis days, more than 90% of the payments were delayed for a certain period of time. This explains the economic significance of the coefficients on these days.

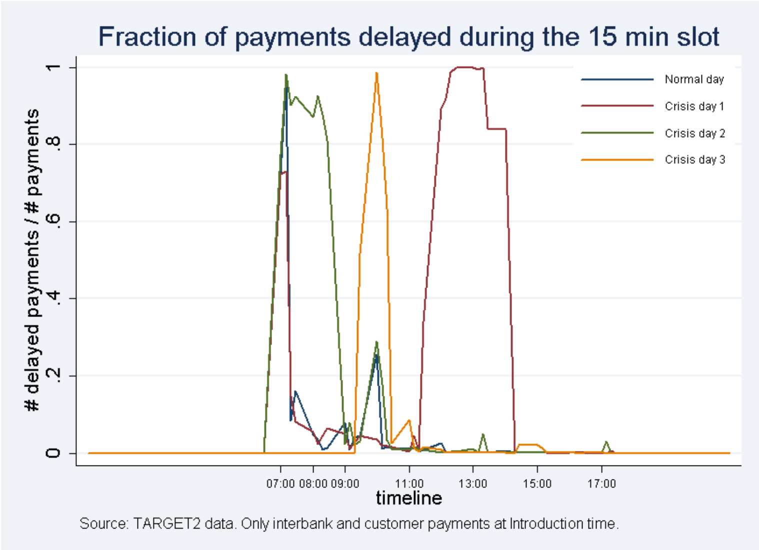


Figure 11: Fraction of payments delayed during the 15-minute slot on the 3 crisis days and 1 normal day.

Lastly, we can see all the results of the regressions for one day, 14/05/2010, in Tables 2 and 3.[12] We notice that our control variables of the payment system environment, number of payments in the system and the ratio of high priority payments, are both significant and of more significant economic size that the delay variables. They also have expected signs; namely; a payment is more likely to be delayed when there are many other payments in the system and more processing time is needed, and they are less likely to be delayed when there are more payments of high priority.

[TABLES WITH REGRESSIONS TO BE INCLUDED HERE]

All in all, it seems that banks do not systematically take strategic bilateral decisions towards other participants on a payment-by-payment basis, as each payment is usually too small to induce a strategic game. They rather indicate that banks make their choices on the liquidity to start operating in the system at the beginning of the day and a mechanical process runs on its own throughout the day after the initial decisions are set, which embeds some persistent prior decisions on liquidity needs and outflows.

5 Robustness checks

We have run several types of robustness checks, and the results are stable to different specifications. First, we tested several variations of dependent variables: a payment is delayed or not and delay value for both reaction and pass-through. Second, we varied definitions of a delayed payment. In the base case, a payment not settled within 5 minutes is considered to be delayed. In robustness

[12]Tables with regression results are available for all the days in the sample and can be provided on demand.

tests, we define a delay as a payment not settled within 1 minute or 10 minutes. Third, we computed our independent variables for rolling windows of different lengths, 30 minutes and 90 minutes, with base definition of 1 hour. Finally, we ran different combinations of independent variables.

6 Conclusion

Recorded payments from a gross settlements system carry information about liquidity in both direct and indirect ways: first, lending and borrowing on the market that is processed solely through this payment system can be inferred through a Furfine algorithm (Furfine [1999]); second, functioning of the payments themselves can be informative about the liquidity on the market. If payments are delayed, then one of the participants, in some sense, "borrow" intraday liquidity from other participants by delaying settlement of payments to them.

We use a unique dataset of transactions from the real-time gross settlement system, TARGET2, that settles the largest payment amounts in the Eurosystem, in order to analyze the behavior of banks with respect to the delay in settlement of interbank and their customers' claims. However, measurement of a liquidity shortfall in a payments system through these delays is difficult due to two reasons: delays propagate to other agents who may also delay, and delays might induce punishing delays on the part of counterparties. We look at the magnitude of these effects and show them to be very small.

We characterize the delays in the TARGET2 system and analyze whether delays in incoming transactions could cause delays in outgoing transactions. To distinguish between the potentially mechanical pass-through of delays and the reaction of one bank to its delaying counterparty, we constructed two sets of instrumental variables and found that the probability of a bank delaying its payments is indeed positively affected by both the likelihood of an incoming payment being late and amount of incoming payments that are delayed. A bank also tends to delay more to its counterparty that delays to it. However, the economic significance of both coefficients is quite small and economically insignificant compared to purely mechanical mechanisms within the gross settlements system. Further, bilateral limits, a mechanism designed to give participants an automatic strategic response to free-riding on liquidity, are not being used by these participants. Intra-day strategic response to liquidity free-riding does not seem to have a great deal of economic significance. We also document that banks do not delay payments based on their transaction value and distributions of value of delayed and non-delayed payments are very similar over the sample period.

Altogether, contrary to the theoretical literature, our findings suggest that banks do not manage their liquidity on a payment-by-payment basis nor do they manage it even on a counterparty-by-counterparty basis. This conclusion opens a venue for a more detailed research based on the hypothesis that banks make their choices on the liquidity in the system at the beginning of the day and a mechanical process runs on its own throughout the day after the initial decisions, which embeds some persistent prior decisions on liquidity needs and outflows.

Table 2: Reaction: Results of regressions for the day 14/05/2014.

The dependent variable is equal to 1 if a payment is delayed and 0 if not. We measure the delay response (Reaction) of bank i in paying bank j at time t_0 as a probability that the delay occurs. It is a linear function of the percentage of value that bank j delayed directly to bank i during the time period $[t_0 - s, t_0)$ just before bank i's payment is made. Bandwidth is equal to 45 for No FE, 8 - Node FE; 5 - Edge FE. t-statistics in parentheses ="* p<0.05; ** p<0.01; *** p<0.001"

	OLS + Hour FE	IV Linear Regression							
		Instrument set 1				Instrument set 2			
		No FE	Hour FE	Node FE	Edge FE	No FE	Hour FE	Node FE	Edge FE
Reaction	-0.000144**	0.00456***	0.00274***	0.00321***	0.00826***	0.00217***	-0.000523**	0.00142***	0.00928***
	(-3.09)	(24.18)	(15.04)	(17.48)	(32.62)	(12.46)	(-3.13)	(8.37)	(29.81)
Incoming Delayed Value	0.000609***	0.00176***	0.000648***	0.00220***	0.00140***	0.00190***	0.000604***	0.00239***	0.00128***
	(11.85)	(34.34)	(12.89)	(38.28)	(22.62)	(37.49)	(12.15)	(41.98)	(19.43)
Total Incoming Value	-0.00183***	-0.00116***	-0.000335*	-0.00593***	-0.00303***	-0.00248***	-0.00202***	-0.00667***	-0.00252***
	(-14.20)	(-6.50)	(-2.14)	(-22.49)	(-9.54)	(-14.38)	(-13.42)	(-25.51)	(-7.59)
Total Outgoing Value	0.00188***	0.00430***	0.00269***	0.00240***	0.00342***	0.00354***	0.00177***	0.00213***	0.00359***
	(14.28)	(26.85)	(19.37)	(12.76)	(16.09)	(22.57)	(13.03)	(11.38)	(16.61)
Number Of Payments In The System	0.00361*	0.0749***	0.00358**	0.0665***	0.0472***	0.0763***	0.00362**	0.0680***	0.0463***
	(2.07)	(71.03)	(3.10)	(64.55)	(43.64)	(73.23)	(3.17)	(66.25)	(42.03)
Ratio Of High Priority Payments	-0.0712***	-0.149***	-0.0728***	-0.170***	-0.142***	-0.155***	-0.0710***	-0.174***	-0.139***
	(-11.68)	(-35.85)	(-18.52)	(-42.45)	(-34.83)	(-37.85)	(-18.25)	(-43.59)	(-33.79)
Constant	0.957***	-0.560***	0.919***			-0.560***	0.919***		
	(58.82)	(-63.45)	(80.60)			(-63.45)	(80.60)		
Observations	190647	190647	190647	190592	180838	190647	190647	190592	180838

Table 3: Path-through: Results of regressions for the day 14/05/2014.

The dependent variable is the amount delayed (in logs) by bank i to bank j. It takes a value 0 if the payment was not delayed and the value of the payment if it was delayed. We measure if bank i delays bigger amount to bank j at time t_0 if it faces larger volume of incoming delayed payments during the time period $[t_0 - s, t_0)$. Bandwidth is equal to 45 for No FE, 8 - Node FE; 5 - Edge FE. t-statistics in parentheses ="* p<0.05; ** p<0.01; *** p<0.001"

	Linear Regression							
	OLS				Instrument set 1			
	No FE	**Hour FE**	**Node FE**	**Edge FE**	**No FE**	**Hour FE**	**Node FE**	**Edge FE**
Incoming Delayed Value	0.0443***	0.0341***	0.0555***	0.0474***	0.0450***	0.0348***	0.0542***	0.0465***
	(24.95)	(29.26)	(69.76)	(58.99)	(62.08)	(40.29)	(67.41)	(57.42)
Total Incoming Value	-0.0809***	-0.0963***	-0.0332***	-0.0319***	-0.0855***	-0.102***	-0.0178***	-0.0219***
	(-26.21)	(-36.08)	(-10.95)	(-9.36)	(-46.29)	(-54.28)	(-5.46)	(-6.05)
Total Outgoing Value	0.151***	0.184***	0.0751***	0.0817***	0.153***	0.187***	0.0741***	0.0809***
	(42.61)	(60.23)	(31.50)	(30.24)	(84.71)	(99.29)	(31.07)	(29.92)
Number Of Payments In The System	1.177***	0.768***	1.091***	1.233***	1.174***	0.761***	1.106***	1.242***
	(48.84)	(21.92)	(151.91)	(168.31)	(185.95)	(80.06)	(152.25)	(167.86)
Ratio Of High Priority Payments	-4.583***	-3.464***	-5.416***	-5.479***	-4.597***	-3.465***	-5.435***	-5.492***
	(-19.60)	(-12.74)	(-91.49)	(-90.83)	(-76.89)	(-49.04)	(-91.78)	(-91.01)
Constant	-9.473***	-5.529***			-9.403***	-5.420***		
	(-41.55)	(-16.58)			(-137.36)	(-56.84)		
Observations	276550	276550	276496	265771	276550	276550	276496	265771

References

P. Abbassi, F. Fecht, and J. Tischer. The intraday interest rate - what's that? *Deutsche Bundesbank Working Paper*, 24, 2015.

L. Bartolini, S. Hilton, and J. McAndrews. Settlement delays in the money market. *Journal of Banking & Finance*, 34(5):934–945, 2010.

M. Bech, J.T.E. Chapman, and R. Garratt. Which bank is the "central" bank? an application of markov theory to the canadian large value transfer system. *Bank of Canada Working Paper*, 2008.

M.L. Bech. Intraday liquidity management: a tale of games banks play. *FRBNY Economic Policy Review*, (Sep):7–23, September 2008.

M.L. Bech and R. Garratt. The intraday liquidity management game. *Journal of Economic Theory*, 109(2):198–219, 2003.

M.L. Bech and K. Soramaki. Liquidity, gridlocks and bank failures in large value payment systems. *E-money and Payment Systems Review*, pages 113–127, 2002.

E. Benos, R. Garratt, and P. Zimmerman. Bank behaviour and risks in chaps following the collapse of lehman brothers. Bank of England working papers 451, Bank of England, June 2012.

W. Beyeler, R.J. Glass, M.L. Bech, and K. Soramäki. Congestion and cascades in payment systems. *Staff Reports, Federal Reserve Bank of New York.*, 259(5), 2006.

L.E. Blume, W.A. Brock, S.N. Durlauf, and Y.M. Ioannides. Identification of social interactions. Economics Series 260, Institute for Advanced Studies, October 2010.

Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150:41–55, 2009.

S. Buckle and E. Campbell. Settlement bank behaviour and throughput rules in an RTGS payment system with collateralised intraday credit. Bank of England working papers 209, Bank of England, December 2003.

M. Diehl. Measuring free riding in large-value payment systems: the case of target2. *Journal of Financial Market Infrastructures*, 1(3):1–23, Spring 2013.

ECB. Target2 annual report. Technical report, European Central Bank, 2013.

J.T. Fox. Estimating matching games with transfers. *NBER Working Paper No. 14382*, 2008.

C. Furfine. The microstructure of the federal funds market. *Financial Markets, Institutions, and Instruments*, 8(5):24–44, 1999.

M. Galbiati and K. Soramaki. An agent-based model of payment systems. *Journal of Economic Dynamics and Control*, 35(6):859–875, 2011.

R. Heijmans and R. Heuver. Is this bank ill? the diagnosis of doctor target2. Technical Report 2(3), Journal of Financial Market Infrastructures, 2014.

M. Manning, E. Nier, and J. Schanz. *The Economics of Large-value Payments and Settlement. Theory and Policy Issues for Central Banks.* Oxford University Press, Great Clarendin Street, Oxford OX2 6DP, 2009.

C.F. Manski. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60:531–542, 1993.

M. Massarenti, S. Petriconi, and J. Lindner. Intraday patterns and timing of target2 interbank payments. *Journal of Financial Market Infrastructures*, 1(2):3–24, Winter 2013.

J.-C. Rochet and J. Tirole. Controlling Risk in Payment Systems. *Journal of Money, Credit and Banking*, 28(4):832–862, November 1996.