Intraday liquidity and the settlement of large-value payments: a simulation-based analysis

Asbjørn Enge, senior advisor, and Frode Øverli, advisor, Payment Systems Department

Interbank systems are of great importance to the economy and the financial system. Using simulations based on real data from Norges Bank’s settlement system, this article illustrates trade-offs between delayed payments and liquidity usage in interbank settlement systems. The simulations demonstrate, for example, that the speed with which payments are settled may be affected by changes in the liquidity available to settlement participants. The effect of optimisation routines in the settlement system is also simulated.

1. Introduction

Banks are linked together by interbank systems, through technical systems and agreements for clearing and settling money transfers between banks. Norwegian interbank systems comprise of several systems with different clearing and settlement procedures for retail payments, securities trading and individual large-value transactions. Gross turnover in the Norwegian Interbank Clearing System (NICS), which is the largest system, is on average NOK 200 billion per day. The bulk of these transactions is settled over banks’ accounts in Norges Bank’s settlement system (NBO). The average daily value of settlements in NBO is over NOK 150 billion. Most large-value payments in NBO are settled in NICS-SWIFT® gross settlements. Chart 1 shows that these settlements also account for the bulk of turnover in NBO.

Settlement systems for large-value payments are central to the financial infrastructure, due to the size of the payment transactions and the fact that it is important that they are executed correctly and at the right time. Smoothly functioning systems for large-value payments are thus crucial to the efficiency of the financial markets, the stability of the financial system and the implementation of monetary policy in a country. As they are typically regarded as systemically important, central banks and supervisory authorities have a particular interest in how these systems are organised and operated (see separate box).

In an efficient payment and settlement system, payments are carried out cost efficiently and with low risk. For participants in the financial sector, the cost of carrying out payment transactions includes the cost of producing payment services, the cost of any payment delays and the cost of payment system participants having to keep a different asset portfolio than they might otherwise have done, in order to execute payments. This may, for example, take the form of deposits in the settlement bank and securities that provide borrowing rights for carrying out settlement.

Berger, Hancock and Marquardt (1996) present a theoretical framework for analysing the trade-off between risk (e.g. delayed completion of payment) and costs in the payment system (e.g. liquidity costs). A payment system is deemed to be technically efficient if costs are minimised at a given risk level and risk is minimised at a given cost level. The simplified illustration given in Chart 2 shows risk (settlement delays) and costs (liquid-
ity usage), where the curve, FF, represents a set of technically efficient points. The curve also shows that risk rises at an increasing rate as costs are reduced (convexity). Innovations in the payment system, for example, technical developments that make it possible to carry out payment faster at a given liquidity level, shift the curve inwards (towards F’F’). Where the outcome on this line occurs depends on the preferences of participants in the payment system, represented by curve II. All points on II are in principle equal for all participants. The curve’s concave form reflects the assumption of a decreasing marginal utility of risk reduction, in other words, that participants are less willing to pay for risk reduction from a starting point of low risk, than for a similar reduction from a high risk level. A number of such curves can be drawn inside and outside II, where participants are more satisfied the closer the curve is to the origin in the chart, i.e. the lower risk and costs are. Point A is the outcome here, given the participants’ trade-off between delays and liquidity usage and the technical possibilities represented by F’F’.

Using a simulation-based approach, this article will illustrate the trade-offs that exist between payment delays and liquidity usage in interbank settlement systems. A number of key concepts and features of settlement systems are introduced in the next section.

2. Features of settlement systems

a) Gross and net settlement

Large-value payments can either be settled individually in gross systems or included in a clearing that is then settled in a net system. Other solutions (hybrid systems) also exist. Gross and net systems entail different risks and costs for settlement participants. Three key risk/cost elements in a settlement system are liquidity, risk of delay and credit risk.

In a net system, participants settle the result from an earlier clearing of incoming and outgoing payments at designated times. Given the interval that elapses between the time that transactions are submitted for clearing and the final settlement of the clearing, banks receiving funds in the settlement implicitly provide credit to other participants for this period. If a bank that owes money in the clearing experiences solvency problems after the transactions have been submitted for clearing, but before final settlement, other banks will be exposed to credit risk in relation to that bank. In this way, the settlement system may cause the spread of solvency problems from one bank to others. This is often called systemic risk and is potentially a danger to the stability of the financial system.3 In gross systems, or RTGS (Real Time Gross Settlement), positions between banks are settled on an individual basis continuously throughout the day, as soon as the payment transaction enters the system. A payment transaction can only be settled if the participant has cover (sufficient liquidity) in their account in the settlement bank. When this account is debited, the payment is completed with final effect. The continuous settlement of transactions entails no credit risk in these systems. Settlement systems for large-value payments have increasingly been based on RTGS (see box).4

From a risk/cost perspective, there are different advantages and disadvantages attached to gross and net settlement systems. Net settlement economises on liquidity, as participants only require the amount needed to cover the results of the clearing. However, as settlement is delayed, net settlement does expose participants to potential credit risk. RTGS settlement is carried out swiftly and does not involve credit risk, but requires more liquidity, as payment transactions are settled individually. Efficient liquidity management throughout the day is therefore important for participants in such settlement systems.

b) Intraday liquidity and transaction cycles

Banks are expected to settle their obligations to customers and other banks in time. They therefore need liquidity, i.e. funds that can be used as means of payment in

---

3 Net systems can, however, be organised in such a way that credit risk is managed, for example, with deferred customer crediting, limits on counterparty exposure (caps), loss sharing agreements, etc.

4 For a more detailed description of RTGS systems, see BIS (1997). For the introduction of RTGS in Norway, see Grønvik and Vedel (1999).

---

"The RTGS revolution"

Net systems with end-of-day settlement were replaced by RTGS systems with continuous settlement throughout the day in a number of countries in the 1990s. Technological developments and central banks’ focus on systemic risk were important reasons for this transition (BIS 2005). In Norway, RTGS was introduced in 1999 in connection with the modernisation of the settlement system in Norges Bank (NBO).
In order to economise on liquidity, banks have to manage their liquidity so that payment obligations can be settled at the right time in the course of the day. If banks carry out settlement in a central bank, intraday liquidity typically takes the form of deposits and borrowing facilities against pledged securities. In NBO, borrowing facilities generally account for the bulk of banks’ disposable liquidity through the day; see Chart 3, which shows disposable liquidity in NBO in the form of borrowing rights and deposits at start-of-day.

As long as the settlement bank does not provide unlimited, unsecured and free credit to banks participating in the settlement, banks will incur costs in connection with acquiring and maintaining liquidity in order to fulfil their payment obligations. These costs are linked to the acquisition of liquid funds (direct funding costs) and also the maintenance of deposits in a settlement account in the central bank (alternative cost in the form of lost interest income). The fact that participating banks have to pledge securities as collateral for borrowing rights in the central bank does entail alternative costs to the extent that it influences banks’ choice of securities portfolio. Banks also incur direct costs in connection with liquidity management.

In addition, costs accrue if settlement is delayed or transactions are not settled at all. As the payments that are transferred in interbank systems are often large or time-critical, the costs incurred for banks and their customers may be substantial if transactions are not settled at the anticipated time. The fact that costs accrue in connection with maintaining liquidity and in the event of delayed settlement, banks have incentives to reduce their liquidity costs, but without it resulting in delays. The trade-off between liquidity costs and the cost of settlement delay is thus an important consideration for banks when adjusting their liquidity levels. Different banks may have different preferences as regards this trade-off, and these can change over time. If a bank’s costs in connection with settlement delays are substantial in relation to liquidity costs (e.g. because many of the transactions are time-critical), it will probably choose to hold more liquidity in order to avoid delays in the settlement.

Several conditions affect participants’ liquidity requirements in a RTGS system. Incoming payments from other settlement participants are one important source of liquidity. The structure of the banking and settlement system and how payments flow through the day will also be of importance to participants’ liquidity requirements (i.e. how evenly incoming and outgoing payments are distributed). A bank can influence its liquidity needs if outgoing payments are managed to coincide with incoming payments. Coordination between settlement system participants may help to reduce the liquidity need and the risk of delays. This can be achieved through the use of shared information systems and the general agreement and regulatory framework for the settlement system in question, including any arrangements for coordinating the exchange of transactions over the course of the day. Such an arrangement may help to prevent situations where individual participants intentionally wait for liquidity from incoming payments before placing their own transactions in the system (free-riding). In order to economise on liquidity usage, banks in Norway have coordinated the exchange of gross transactions in NBO.

A well-designed settlement system can help to make liquidity usage more efficient, which is particularly pertinent in situations where there is not much liquidity and payments are queued. If the settlement system includes elements of both gross and net settlement, improved recycling of liquidity can be achieved. In RTGS systems, for example, it is usual to have queue management mechanisms, where transactions that do not have sufficient liquidity to be settled are placed in a queue in the settlement system. These transactions then await new liquidity from later payments and are settled according to more detailed rules for settlement prioritisation and sequencing. Systems may also include netting procedures for transactions in the queue, where the netting effect is bilateral between two participants and/or multilateral between several participants. Features such as queue management and netting procedures minimise participants’ liquidity need at the same time as reducing the delays in settlement. The result is a better trade-off between liquidity and settlement speed than might otherwise be achieved in a purely gross or net settlement.

The article will use a simulation-based approach to illustrate trade-offs between different levels of bank liquidity and payment delays in NBO settlement. But first of all, the data and methodology on which the analysis is based will be presented.

---

5 For a discussion on payment flows and ways in which to influence transaction patterns, see Trundle & McAndrews (2001), pp. 131–133.
3. Data and method

A simulation tool developed by the Finnish central bank makes it possible to carry out simulations based on actual settlement data. The simulator can be used to analyse the effect of changes in the liquidity available to settlement participants and/or the introduction of new settlement procedures. The effect on variables such as the liquidity requirements, payment delays and the settlement ratio can then be studied.

The simulations presented in this article were carried out using settlement data from the RTGS system in NBO (NICS-SWIFT gross transactions) and by generating systems data in the simulator. The relevant data from NBO includes participating banks, transactions between participants (time, sender/receiver and amount) and their available liquidity (balance in settlement accounts and borrowing facilities). The settlement procedures and rules are defined in the simulating tool, including the system’s opening hours, how transactions are settled and any optimisation routines (queuing function, netting procedures, etc.).

The analysis is based on settlement data from 10 days in October 2005. The days can be characterised as relatively normal, with transactions for a value of NOK 160 billion on average being settled in NICS-SWIFT gross settlements per day. This accounts for around 87 per cent of total turnover in NBO in the period. On average, there were 558 transactions per day and the average size of transactions was NOK 287 million. A maximum of 20 banks participated in the settlement. The settlement volume is relatively concentrated. The five largest banks accounted for over 88 per cent of the transaction value. As the banks coordinate the exchange of transactions, the bulk of the settlement is carried out between 12.30pm and 1.30pm (69 per cent of the turnover value). When presenting the results, the average for the days in the period has been used.

Two types of simulation have been carried out. In section 4 a) the theoretical reference points for the amount of liquidity needed to finalise a given flow of gross transactions are calculated. In sections 4 b) and c) the effects of varying liquidity on payment delays and the volume of unsettled transactions are studied. NBO’s features have been imitated as closely as possible here, based on the following settlement procedures: when a settlement participant places a payment transaction in the system, the transaction will be settled immediately if there is sufficient liquidity (a positive balance and/or borrowing rights). If the participant lacks liquidity, the transaction is placed in a queue until there is sufficient cover for settlement. The transaction will be settled if and when incoming transactions from other participants can supply sufficient new liquidity or if the participant’s borrowing facilities are increased. Transactions that are waiting in a queue are managed according to the FIFO principle (“first in, first out”), which means that transactions are released from the queue for settlement in the order in which they joined the queue (“longest in, processed first”). When a queue starts to form, a gridlock mechanism will also try to offset the transactions between participants both bilaterally and multilaterally. If a participant still lacks liquidity at end-of-day, the transaction will not be settled that day.

In the simulations in sections 4 b) and c) information about participating banks’ actual liquidity in NBO has been used. The simulations were made by changing the level of available liquidity, by adjusting participants’ balances and borrowing facilities by the same percentage. What determines participants’ liquidity needs is the actual transaction flows through the day and the settlement system properties. This entails an assumption that banks’ behaviour remains unchanged even though their liquidity level varies. There is, however, reason to assume that participants’ transaction patterns will also change when liquidity in the settlement is changed. Furthermore, liquidity is in practice not just used for NICS-SWIFT gross settlement, but also for other settlements in NBO. On any given day, other settlements will thus be able to supply or draw in liquidity for the participant in question. The simulations do not take account of the fact that some transactions are time-critical. The results must be evaluated in light of this.

4. Simulation results

a) Theoretical reference points for liquidity requirements

Liquidity requirements in a RTGS system will, among other things, depend on whether payment transactions are settled immediately or whether they are placed in a queue for settlement later. This means that a trade-off between liquidity need and settlement delays has to be considered. This trade-off can be illustrated by calculating the reference points for liquidity needs. The concepts of upper bound (UB) and lower bound (LB) for liquidity requirements are relevant in this context (Koponen & Soramäki, 1998). UB shows how much liquidity a participant in a RTGS system needs to ensure that all outgoing payments are settled immediately when they enter the system (without waiting in the queue). LB shows the minimum amount of liquidity a bank needs to cover its net obligations at end-of-day, for all its transactions through the day. When assessing the trade-off between liquidity needs and the speed with which pay-

---

6 For a description of the simulator (“BoF-PSS2”), see Leinonen & Soramäki (2003) or Bank of Finland: http://www.bof.fi/eng/3_rahoitustoiminta/3.4_Maksujarjestelmat/3.4.3_Kehittaminen/ 3.4.3.3_Bof-ps2/

7 This procedure has been simplified somewhat in the simulations in relation to the actual system.

8 In the simulations, payments in connection with the foreign exchange settlement system, CLS, are treated such that payments from CLS are executed irrespective of banks’ pay-ins. Payments in connection with Scandinavian Cash Pool (SCP) are not included in the simulations.

9 In situations where liquidity is below LB, transactions will remain unsettled, whereas a liquidity level below UB means that payments cannot be settled immediately and have to be placed in a queue.
Economic Bulletin 1/06

If participants had had unlimited borrowing rights, the average liquidity requirement would be 27 per cent of the turnover value (UB). As the table indicates, there is some variation in liquidity requirements over the period.

It is important to emphasise that these limits are theoretical measures. In order for a liquidity level equivalent to UB to actually result in maximised settlement speed, it is assumed, among other things, that the liquidity in the system is optimally divided between participants at all times. No consideration is taken of the fact that a number of transactions may be time-critical or that it is possible for participants to reprioritise transactions that are placed in a queue. In the event of time-critical transactions, LB will, for example, be too low as participating banks have to secure liquidity in order to carry out such transactions at a given time. 10

b) Liquidity and settlement delays

The starting point for the following simulations was to study how changes in actual liquidity available to settlement participants influence the speed with which payments can be settled, or if they can be settled at all. The settlement procedures imitate NBO as closely as possible (cf. section 3). Transaction flows are the same as actual NICS-SWIFT gross settlement in NBO.

Settlement delays as a result of varying liquidity availability can be measured in several ways. One way to express the overall level of delays is with an indicator introduced by Leinonen & Soramäki (1999). The level of delays is measured by ρ:

\[ \rho \]

\[ \rho \] can assume values between 0 and 1. k is an index for each separate payment, s is the value of each payment, t′ is point at which the payment enters the system, t the point at which the payment is settled and T is end-of-day. The indicator is based on the time each separate payment spends in the queue compared with its maximum possible time in the queue. This gives a value-weighted average of the relative delay for all payments. If all payments are settled immediately on entry into the system, \( \rho = 0 \). If \( \rho = 1 \), all payments have remained in the queue from the time they were placed in the system until end-of-day.

Chart 5 shows the effect of changing the level of available liquidity on the extent of delayed settlement, measured by the indicator \( \rho \). A liquidity level of say 40 per

### Table 1

| Liquidity requirement limits as a percentage of total turnover |
|---------------------------------|----------------|----------------|
|                                | Lower bound | Upper bound   |
| Average                        | 5%           | 27%           |
| Highest value                  | 14%          | 33%           |
| Lowest value                   | 2%           | 23%           |

Source: Norges Bank

10 For a more detailed discussion of upper and lower bounds, see Koponen & Soramäki (1998).
cent means that participants are allocated liquidity equivalent to 40 per cent of their actual balance and borrowing rights. The chart shows that generally liquidity must be reduced substantially for the value of the indicator $p$ to rise noticeably. At a liquidity level of 50 per cent, $p$ has a value of 0.05. If the level is reduced to 20 per cent, $p$ more than doubles to 0.11. With a 5 per cent liquidity level, $p$ increases markedly to 0.34.

One key observation is the shape of the curve. The curve is generally convex, which means that the more the settlement participants’ liquidity is reduced, the greater the delay in settlement. Or, in other words, at low liquidity levels, a small injection of liquidity can substantially reduce delays.

Table 2 shows the effects of varying liquidity on some other indicators for delays and for settlement in NICS-SWIFT gross settlement. When available liquidity was halved, the average settlement time was, for example, four minutes.\textsuperscript{11} If liquidity was reduced to 20 per cent, the average settlement time was 19 minutes. The simulation results also show that the value of transactions that remained unsettled at this liquidity level was NOK 7 billion. With a liquidity level of 10 per cent, this figure rose to NOK 16 billion.

\textbf{c) Effect of optimisation features}

NBO contains features for managing queues and gridlock situations. The gridlock mechanism attempts, as the name indicates, to resolve gridlocks, in other words, situations where there is little liquidity and the queue formation means that several payment transactions between banks are awaiting settlement. None of the transactions in the queue can be settled if they are viewed in isolation. If several transactions are taken for settlement at the same time, they could, however, be settled. Thus the gridlock mechanism makes the use of liquidity more efficient by netting the transactions that are waiting in the queue. The netting can be both bilateral (between two participants) and multilateral (between several participants). The purpose is to reduce delays in settlement and the number of transactions that cannot be settled at end-of-day. If there is still insufficient liquidity to settle the payments after this procedure, the situation is characterised as ‘deadlock’. Only the supply of new liquidity will then be able to prevent transactions from remaining unsettled.

Simulations can be used to illustrate the effect of such optimisation routines. As in the last section, the simulations are based on a RTGS system with a FIFO queuing function where a gridlock mechanism attempts to achieve netting effects between participants. Two reference scenarios are made. In the first, a pure RTGS system without queuing and gridlock functions is simulated. In the second, a RTGS system with a FIFO queuing function, but no gridlock mechanism, is simulated. The effect on the settlement ratio in the different scenarios is then compared.

Chart 6 shows the value of total unsettled payment

\textbf{Table 2 Selected indicators at different liquidity levels}\textsuperscript{1}

<table>
<thead>
<tr>
<th>Liquidity level in % of actual liquidity</th>
<th>Settlement delay indicator ($p$)</th>
<th>Average settlement time (min)</th>
<th>Value of unsettled payments (NOK billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>0.34</td>
<td>107</td>
<td>43</td>
</tr>
<tr>
<td>10%</td>
<td>0.20</td>
<td>56</td>
<td>16</td>
</tr>
<tr>
<td>20%</td>
<td>0.11</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>30%</td>
<td>0.08</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>40%</td>
<td>0.06</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>50%</td>
<td>0.05</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>100%</td>
<td>0.02</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Memorandum:

Average daily turnover value in period: NOK 160 billion

\textsuperscript{1} Results are presented as an average of daily observations in the sample that describes the settlement.

Source: Norges Bank

\textsuperscript{11} Defined as the average time a monetary unit has to wait in queue before settlement (time and value-weighted average).
transactions as a share of total turnover. The results show that optimisation routines ensure a considerably higher settlement ratio than in a system without such features. The difference is relatively small between RTGS with a queuing function and RTGS with a queuing and gridlock function at liquidity levels down to 20 per cent. But below this level, the difference increases. One reason that the gridlock mechanism is efficient is that a substantial share of transactions are between a small number of larger participants and bilateral netting effects can thus be achieved. At very low liquidity levels, however, the settlement ratio is reduced noticeably even when there is a gridlock function. This is because the system increasingly experiences “deadlocks”, in other words, only the supply of new liquidity will increase the level of settlement. At a 5 per cent liquidity level, more than 27 per cent of the transaction value was unsettled at end-of-day.

5. Conclusion

Using a simulation-based approach, this article has illustrated relationships between settlement delays and liquidity usage.

The banks that participate in NBO generally hold liquidity levels that entail little delay in payment settlements. The simulations indicate that liquidity must be reduced substantially before considerable settlement delays occur. It must be emphasised that the analysis is based on data from a period with relatively normal transaction volumes and liquidity levels. However, even though the level of liquidity is sufficient in normal situations, the extent of delays and unsettled transactions may become significant when a “critical” liquidity level has been reached. The simulations regarding the effect of optimisation routines show that these do contribute to a higher payment settlement ratio.

The liquidity levels that participants in NBO have chosen, may indicate that the costs of delays in the Norwegian settlement system are deemed to be relatively high compared with liquidity costs. If the relative costs of liquidity increase, banks might adapt to new levels of liquidity and/or adjust their transaction pattern. However, further analyses would be needed in order to establish whether this might result in a greater number of delays and unsettled transactions.

References


