

LIQUIDITY RISK IN SECURITIES SETTLEMENT

Johan Devriese

National Bank of Belgium*

Janet Mitchell

National Bank of Belgium**

This draft: June, 2005

Abstract

This paper studies the potential impact on securities settlement systems (SSSs) of a major market disruption, caused by the default of the largest player. A multi-period, multi-security model with intraday credit is used to simulate direct and second-round settlement failures triggered by the default, as well as the dynamics of settlement failures, arising from a lag in settlement relative to the date of trades. The effects of the defaulter's net trade position, the numbers of securities and participants in the market, and participants' trading behavior are also analyzed.

We show that in SSSs – contrary to payment systems – large and persistent settlement failures are possible even when ample liquidity is provided. Central bank liquidity support to SSSs thus cannot eliminate settlement failures due to major market disruptions. This is due to the fact that securities transactions involve a cash leg and a securities leg, and liquidity can affect only the cash side of a transaction. Whereas a broad program of securities borrowing and lending might help, it is precisely during periods of market disruption that participants will be least willing to lend securities.

Settlement failures can continue to occur beyond the period corresponding to the lag in settlement. This is due to the fact that, upon observation of a default, market participants must form expectations about the impact of the default, and these expectations affect current trading behavior. If, ex post, fewer of the previous trades settle than expected, new settlement failures will occur. This result has interesting implications for financial stability. On the one hand, conservative reactions by market participants to a default – for example by limiting the volume of trades – can result in a more rapid return of the settlement system to a normal level of efficiency. On the other hand, limitation of trading by market participants can reduce market liquidity, which may have a negative impact on financial stability.

Keywords: Securities settlement, liquidity risk, contagion;
JEL-classification numbers: G20, G28

*E-mail: johan.devriese@nbb.be. **E-mail: Janet.Mitchell@nbb.be. The views and findings of this paper are entirely those of the authors and do not necessarily represent the views of the National Bank of Belgium.

We would like to thank Hans Degryse, Mathias Dewatripont, Stan Maes, and Thierry Timmermans for helpful comments.

1 INTRODUCTION

A prerequisite for the development of a viable capital market is a well functioning transactions infrastructure. The settlement of securities transactions is an important component of this infrastructure, as it determines the legal transfer of the securities that are traded. This infrastructure must operate in a seamless and integrated manner, in order to minimize the costs and risks for the end users in the market and to facilitate the allocation of capital. Hence, securities settlement systems (SSSs) are crucial to the financial system and are often supported by the central bank as lender of last resort.¹

Disruptions in the settlement infrastructure can lead to increased transaction costs and to a possible erosion of market liquidity which, if serious enough, may undermine financial stability. An extreme example of the potential severity of settlement failures was provided by the September 11 attacks. Settlement failures in the U.S. Treasury market jumped from \$1.7 billion per day in the week ending September 5 to \$190 billion per day in the week ending September 19 (see Fleming and Garbade, 2002). Failures rose initially because of the destruction of communication facilities, but remained high because the lending and borrowing program was ill-suited to absorb the massive shock.

This paper explores the potential consequences of a market disruption that is less severe than the Sept. 11 attack but that is nevertheless serious; namely, the default of the largest participant in the market. This type of shock is of interest for policy makers and SSSs alike. Indeed, among the recommendations for securities settlement systems recently set forth by the Committee on Payment and Settlement Systems and the International Organization of Securities Commissions is that "SSSs that extend intraday credit to participants...should institute risk controls that, at a minimum, ensure timely settlement in the event that the participant with the largest payment obligation is unable to settle." (See BIS, 2001). Similar scenarios are also used by SSSs for stress testing purposes.

Although a number of previous studies have analyzed the impacts of major disruptions to payment systems and the extent of resulting contagion, very little investigation of disruptions in SSSs has been undertaken. Such analysis may yield new insights, due to several differences between securities settlement and payment systems which could potentially lead to important differences in the impacts of shocks in the two systems.

A first key difference between payment systems and SSSs is that unlike payments, securities transactions involve a securities leg as well as a cash leg. This gives rise in securities settlement to principal risk, which is the risk that the seller of a security delivers the security but does not receive cash in return or that the buyer of a security makes the payment but does not receive delivery of the security. The response to this risk has been to implement delivery-versus-payment (DVP)

¹ We use the term SSS to refer to all of the participants as well as the financial infrastructure involved in securities settlement.

systems – by which settlement finality of the securities and cash leg occurs at the same time, and thus principal risk is eliminated. However, default by a major participant can still have an impact on liquidity in the SSS if unsettled trades of the defaulted participant are deleted from the system, leaving non-defaulting participants with unanticipated cash or securities positions.

Yet, contrary to payment systems, a disruption in securities settlement cannot be fully accommodated by providing liquidity. This is because after the initial default, participants may not only be short in cash, but also in securities. In order to further eliminate the effects of a settlement disruption, it would also be necessary to have a broad, well functioning program of securities borrowing and lending. Yet, it is precisely during crisis periods that uncertainty about repayment is greatest and holders of securities will be the least willing to lend. As we discuss below, central banks may be able to take some measures to help resolve the shortage of securities; however, these policies have their limits, especially in the case of SSSs that settle non-treasury securities.

A second difference between SSSs and payment systems is the presence of a settlement lag in SSSs. For example, settlement at time $t+2$ implies that the settlement of trades takes place two days after the trades have occurred. Although this lag gives participants extra time to find the necessary funds to finance the trades, it also increases replacement cost risk.² In the delay between trade and settlement, asset prices may have changed, making it possibly more expensive to trade the securities elsewhere if the initial trade does not settle.

More important for this paper, the existence of a settlement lag suggests that a disruption in the settlement system may have impacts lasting longer than a single day. Indeed, the direct effect of a default by a major participant will continue to be observed for the number of days corresponding to the lag in settlement. In addition, the total disruption – which includes the indirect, or contagion, effects of default – may last even longer than the period of the lag. The reason is that although participants are assumed to know their counterparties and, thus, can calculate the direct effect of default, participants do not know the counterparties of their counterparties and cannot know which of the nondefaulting counterparties traded with the defaulting participant and thus will be unable to settle another trade as a consequence. Participants must form expectations about the indirect effects of default, and these expectations determine the quantities of securities and cash upon which they base their trades after the default. If, ex post, actual settlement failures due to the default turn out to be higher than participants had expected, then additional settlement failures may occur.

The differences in payment and settlement systems give rise to a number of important questions for SSSs which have not been addressed by existing literature. What are the dynamic effects on settlement – both direct and contagion effects – of a major disruption in the market? Is the first-day impact greater or smaller than the impact in subsequent days? How many days does it take for

settlement efficiency (the percentage of trades settled) to return to its normal level? How does the existence of a cash leg and a securities leg influence the degree of contagion, relative to a payment system where only cash is involved? Can central bank support of the SSS through credit provision prevent contagion? If so, how much credit would be needed? How does the trade position of the defaulter (e.g., size, net buyer versus net seller) affect the magnitude of the disruption?

This paper uses a simulation to address these questions. Settlement is assumed to occur in a DVP system with gross (trade-by-trade) settlement and with a two-day lag. The SSS may provide liquidity in the form of credit, and results are compared across scenarios with differing assumptions regarding the amount of liquidity provided. Default by the largest player triggers the initial settlement failures. The direct and contagion effects of the default are measured over a period of ten days following the default. The impact of the defaulter's net trade position, as well as the numbers of securities and participants in the market, and participants' trading behavior (moderate versus more extreme) are analyzed.

Several results emerge from the analysis. First, the two-day lag in settlement implies that settlement failures will last for at least two days following default. Settlement efficiency is in fact lower on the day following default than on the day that default occurs, due to continuing contagion.³ Thus, the crisis situation initially worsens before improving. Interestingly, settlement efficiency may not return to normal after two days (and indeed does not return to normal in the simulations reported here), despite the two-day settlement lag. The reason is that, as noted above, upon observation of a default on day t , market participants must form expectations about the impact that the default will have on their unsettled trades. If, ex post, fewer of the unsettled trades actually settle than anticipated, then participants may commit to trades in the two days following the default that later turn out to be infeasible.

This result has interesting implications for financial stability. On the one hand, conservative reactions by market participants to a default – for example by limiting the volume of trades – will result in a more rapid return of the SSS to a normal level of efficiency, and an end to the crisis. On the other hand, limitation of trading by market participants can reduce market liquidity, which may have a significant, negative impact on financial stability. In addition, limitation of trading can have negative welfare effects on participants, due to lost benefits from trading.

A second result is that the net trade position of the defaulting institution can have a significant impact on the severity of the crisis. When the SSS provides little or no liquidity, a large net buy position of the defaulter will cause a significantly higher fall in settlement efficiency than will a large net sell position. Generous liquidity provision by the SSS can eliminate the differential effects of the

² Replacement cost risk is the risk that a counterparty may default prior to settlement, denying the non-defaulting party the gain on the transaction (BIS, 1992).

³ Default during day t , prior to settlement on that day, implies that the trades from day $t-2$ are the first to be affected, as these trades are settled at the end of day t . Settlement on the day following default concerns trades that occurred on day $t-1$.

defaulter's trade position on settlement efficiency; however, liquidity provision cannot completely eliminate settlement failures: settlement efficiency still falls significantly following the default of the largest player even when plenty of liquidity is available.⁴ As suggested above, this is due to the fact that liquidity provision by a central bank or a central security depository (CSD) can eliminate problems on the cash side of transactions but not on the securities side. Thus, default by a major player can still have an impact on the system.

Additional results concern the implications of differing numbers of participants and differences in trading behavior. Not surprisingly, the severity of the crisis (in terms of settlement inefficiency) decreases as the number of participants increases. The size of the largest participant relative to the market (or the degree of concentration) is linked to the number of participants; hence, a larger number of participants translates into a smaller direct impact of default by the largest player. In contrast, the severity of the crisis increases with aggregate trading volume. The crisis is also more severe when participants' trading behavior causes them to trade closer to the boundaries of their budget constraints (more "extreme" trades) than when their trades are more "moderate".

The paper proceeds as follows. Section 2 reviews the existing literature. Section 3 presents a stylized example of securities trading, which illustrates some of the main ideas and effects of the simulation model. Section 4 presents the model used for the simulations. Section 5 discusses the simulation results. Section 6 concludes.

2 LITERATURE REVIEW

Contagion has become a topic much investigated in the finance literature during the last decade. Starting with bank runs as a channel of financial contagion when agents do not have complete information (Diamond and Dybvig, 1983), authors have shown that even under perfect information financial contagion is possible (see, e.g., Rochet and Tirole (1996), Allen and Gale (2000) and Diamond and Rajan (2003)). These papers concentrate on contagion in the interbank market, where it is assumed that banks have uncollateralized exposures to each other, and the default of one bank can cause other banks to become insolvent and to default as well. Hence, credit risk and solvency risk are at the fore.

This is also the idea behind several empirical studies investigating financial contagion. Humphrey (1986), Angilini, Maresca and Russo (1996) and Northcott (2002) all use payments data from a single day in payments systems in which net settlement occurs. Humphrey uses data from a randomly selected business day in CHIPS (U.S.) and simulates the impact of a major participant's failure by unwinding all of the day's transactions to and from that participant, calculating the balances of the remaining participants, comparing this with their capital buffers, and iterating the unwind. Humphrey finds that on average, 37% of the institutions fail after the initial participant's failure. Angilini, Maresca, and Russo (1996) and Northcott (2002) use a similar method for the

⁴ Indeed, the impact of additional liquidity provision above some threshold appears to be limited.

Italian and Canadian netting systems, respectively, and conclude that systemic risk in those systems is very low or nonexistent.

Other empirical studies use data on interbank exposures as reported in banks' balance sheets (see Upper and Worms (2002), Furfine (2003), and Degryse and Nguyen (2004)). On average, these papers also find low degrees of potential contagion. In reaction to these findings, however, Cifuentes, Ferruci, and Shin (2004) argue that in reality systemic risk may be significantly greater than that identified by the interbank contagion simulations, because market risk may materialize in addition to the credit risk. That is, if following a default by an interbank borrower, bank creditors must liquidate collateral in order to meet their own interbank obligations, then asset prices may fall, thereby lowering banks' values even further and possibly generating new defaults. Cifuentes et al illustrate the potential impact of market risk via simulations where bank assets are marked to market and banks' sales of illiquid assets in response to defaults by interbank borrowers are assumed to cause a fall in the market prices of these assets.

The small degree of interbank contagion found empirically, together with the virtual absence of principal risk (or credit risk) in SSSs may explain why there are very few studies on systemic risk in securities settlement systems. Indeed, De Bandt and Hartmann (2000) observe in their extensive literature review on systemic risk that: "Empirical studies of systemic risk in securities settlement systems appear to be non-existent". However, as noted in the Introduction, liquidity risk is important in SSSs and may have systemic consequences.⁵

The nature of systemic risk in SSSs, however, seems to depend somewhat upon the type of settlement system: net or gross settlement. In net settlement systems, transactions are settled on a net basis, which economizes on the amount of liquidity needed by participants. However, default by a participant in a net settlement system causes costly trade unwinds, whereby some or all of the transfers involving that participant are deleted and the settlement obligations of the other participants are recalculated, which may lead to possible further unwinds. This increases replacement cost risk, as settlement is only final at the end of the entire settlement process.

Gross settlement systems, on the other hand, transfer instructions for both securities and funds on a trade-by-trade basis during the settlement process. Failure of a participant to meet a delivery or payment obligation on a given transaction will not lead to costly unwinds of multiple transactions. Yet, DVP systems with gross settlement require substantial intraday liquidity. If participants are unable to adjust their cash balances during the processing cycle, they will have to hold enough cash to cover at least the largest debit position during processing. Hence, liquidity risk becomes more important than with net settlement systems. If sufficient money balances are not available,

⁵ It is possible that the liquidity risk arising from disruptions in SSSs can also lead to market risk, if participants liquidate collateral in response to cash shortages arising from settlement failures. However, the consequences of such market risk would be, as in the model of Cifuentes et al, to lower participants' equity values and possibly to cause insolvencies. Whereas defaults due to a fall in asset prices would lead to additional liquidity

high “fail” rates may result, implying substantial liquidity risk and replacement cost risk to counterparties.

Most of the literature on payments settlement has focused on the differential effects of gross versus net settlement. Angilini (1998) uses a real-time gross settlement model (RTGS) to show that if daylight liquidity is costly, banks may be induced to postpone payment, hence increasing liquidity risk in the system. On the other hand, Kahn and Roberds (1998) note that although net settlement is less costly due to the lower need for liquidity, net settlement increases moral hazard, as banks have an option to revoke their trades, which distorts incentives. Kahn, McAndrews and Roberds (2003) analyze more fully the prospect of strategic default in settlement systems and end up concluding that net settlement causes less payment gridlock. Leinonen and Soramaki (1999) use Finnish data to quantify the relationship between liquidity usage and settlement delay in net settlement systems and in RTGS systems with queuing of unsettled trades. When the central bank provides low-cost intraday credit, liquidity costs are low relative to delay costs; hence, RTGS systems with queuing are more efficient.

The main conclusion from this line of research appears to be that there is no liquidity risk in payment systems using gross settlement as long as there is sufficient and cheap intraday credit. However, for the reasons noted in the Introduction (e.g., securities and cash leg, settlement lags), this argument will not necessarily hold for SSSs. The only paper to our knowledge that investigates liquidity risk in SSSs is that of Iori (2004), which analyzes the importance of operational risk with respect to differing lag times between trade and settlement in both gross and net settlement systems. In this model, only one security is traded, and no cash or budget constraints exist for the participants in the system. Trades occur at periodic intervals, and operational delays result in settlement failures whenever the operational delay is longer than the lag between trade and settlement. While shortening the lag between trade and settlement has the advantage of reducing replacement cost following the failure of a participant to settle, it also increases the likelihood of settlement failures caused by an operational problem. Thus, even under real-time settlement ($t+0$), significant settlement contagion is still possible.

Much of the empirical literature on contagion in payments systems and interbank markets makes use of simulations with strong underlying assumptions, which are necessary because of the inability to obtain data on participants' bilateral positions. This will be all the more true for simulations of SSSs, which generally will not be able to make use of any real data. Not only are data relating to individual trades in SSSs highly confidential, but also the amount of data needed for an empirical study would be massive, due to the need to have data on participants' cash and securities holdings as well as their trades. Only the SSSs themselves are able to use real data in simulations or stress tests. Unfortunately, such exercises are for internal use only and often suffer from a number of shortcomings when viewed from a financial stability perspective. First, SSSs are mainly concerned about their own exposure in case the largest participant fails. Second, stress

risk for the nondefaulting participants, there is no direct feedback mechanism from the market risk to liquidity

tests often take into account only the direct effects of a participant's failure, which underestimates systemic risk. Moreover, when contagion, or second-round, effects are incorporated, they only cover a single day of trade data, while disruptions in the settlement system may last for several days. Finally, trading behavior in times of stress likely differs significantly from behavior on "normal" days, raising the question as to whether the use of trading data from a "normal" day is valid for simulating a stress event. The only apparent way around this problem is to conduct empirical studies based upon real stress events, as in Fleming and Garbade (2002). (Un)fortunately, these events are rare.

3 STYLISTED EXAMPLE

This section presents a stylized example of securities trading, which illustrates settlement with a DVP system, the effect of a settlement lag, and the liquidity risk arising from the default of a participant. The example is one where a single security is traded and where there are four participants (W, X, Y, and Z) in the system. Settlement occurs with a two-day lag. Trades from day D are assumed to be settled at the end of day D+2 (after trading on day D+2 has already occurred).

The diagram below depicts both trading and settlement. Day D corresponds to the day on which Participant W is assumed to default. The diagram illustrates trading between all of the participants from day D-2 through day D+1 and settlement of trades on day D through day D+3 (the two-day lag implies that trades from day D-2 will not be settled until day D).

The left panel for each day relates to trading and displays in parentheses the quantities of the security and cash, respectively, that each participant expects to have after incorporating past trades into initial endowments.⁶ Trades are represented by arrows, where the direction of the arrow indicates who is selling the security to whom, and the number without parentheses next to each arrow represents the quantity of the security traded. The number in parentheses next to each arrow represents the order in which the trades occur. All securities prices are assumed equal to 1; therefore, a sale of 30 securities by W to X (the first trade on day D-2) will also involve a payment of 30 in cash by X to W. Trades are constrained to be feasible given the participants' expected holdings of securities and cash, and no short selling or credit is allowed. Participants' expected holdings of the security and cash are updated after each trade and used to determine the feasibility of the subsequent trade.

The right panel for each day relates to settlement, with participants' actual holdings of the security and cash just prior to the beginning of the settlement process (and used for settlement) given in parentheses for each participant. Trades cannot be partially settled; i.e., either the entire trade is settled or no quantities are exchanged. Trades settle when both participants hold in their accounts

risk.

⁶ The quantities of the security and cash given for each participant on day D-2 can be taken as the initial endowments.

the necessary quantities of the security or cash required by the trade. If one of the participants is short, the trade does not settle (DVP system). Gross settlement is used in the example; therefore, netting (i.e. offsetting buy and sell positions) is not allowed. Trades that settle are represented by a solid arrow. Trades that do not settle are represented by a dashed arrow. Settlement of trades is attempted in the order in which the trades occurred (which maximizes settlement efficiency).

Panel a of the diagram (day D-2) gives the initial endowments of the security and cash for each participant. For example, X starts with 100 units of the security and 50 cash. In the first trade, W sells 30 of the security to X, who now expects to have 130 of the security and 20 cash. This allows X to sell 110 of the security to Y in the second trade. No settlement occurs on day D-2 because of the two-day settlement lag.

On the next day (day D-1), participants' expected holdings of the security and cash incorporate all of the trades from the previous day (panel b). For instance, participant X now expects to have 60 of the security ($100+30-110+40 = 60$) and 90 in cash ($50-30+110-40 = 90$). In this example, W engages in no trading on day D-1.⁷

We assume that at the beginning of day D, and before any trading has begun, participant W defaults. All other participants know that all unsettled trades with W will be deleted from the system. They must, therefore, adjust their expected quantities of the security and cash accordingly. This is the direct effect of the default. For example, participant X knows that it will not receive 30 of the security from W. Consequently, X now has 30 less of the security and 30 more in cash than it would have expected to have if the default by W had not occurred.

As mentioned above, whereas we assume that participants know the counterparties of their trades, they are not assumed to know the quantities of the security and cash held by their counterparties, nor the trades of their counterparties with other counterparties. Hence, participants cannot anticipate which of their previous trades with nondefaulting counterparties will not settle due to the impact of W's default on those counterparties. In other words, participants do not know the indirect effects of the default. As a consequence, they must form some expectations about these effects. To keep this example as simple as possible, we make the (unrealistic) assumption that participants expect the indirect effects of default to be zero; that is, they expect all trades with nondefaulting counterparties to settle.⁸ For example, Y's expected holdings, which will serve as the basis for trading on day D (panel c), become 10 of the security ($95-90+5 = 10$) and 160 in cash ($75+90-5 = 160$).

At the end of day D, trades from day D-2 are presented for settlement. Panel d shows that trades from W have been deleted from the system (hence there is no arrow corresponding to the trade between W and X on day D-2). Because this trade has been deleted (and X does not receive the

⁷ We make this assumption only to keep the example as simple as possible.

⁸ We adopt a richer assumption on expectations in the model used for our simulations.

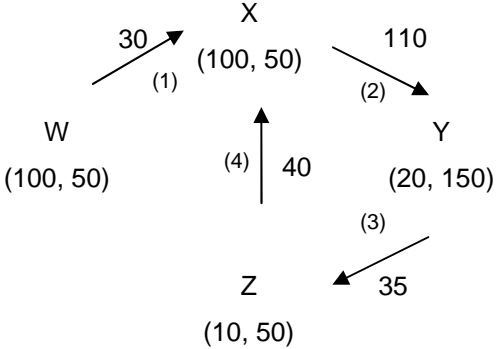
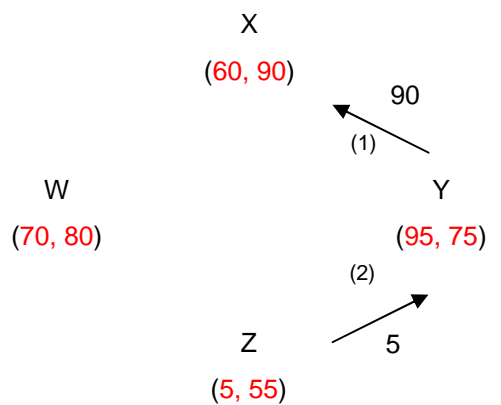
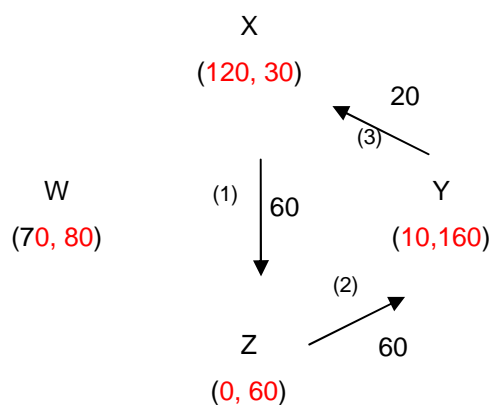
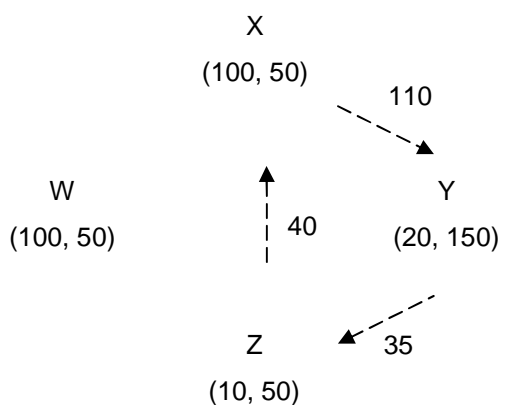
30 in securities from W), the trade between X and Y (and indeed, each subsequent trade) does not settle (dashed arrows). These are the indirect effects of the default. X, Y and Z are all securities-constrained as a result. These unsettled trades are put in the queue of unsettled trades and will be presented again for settlement on Day D+1.

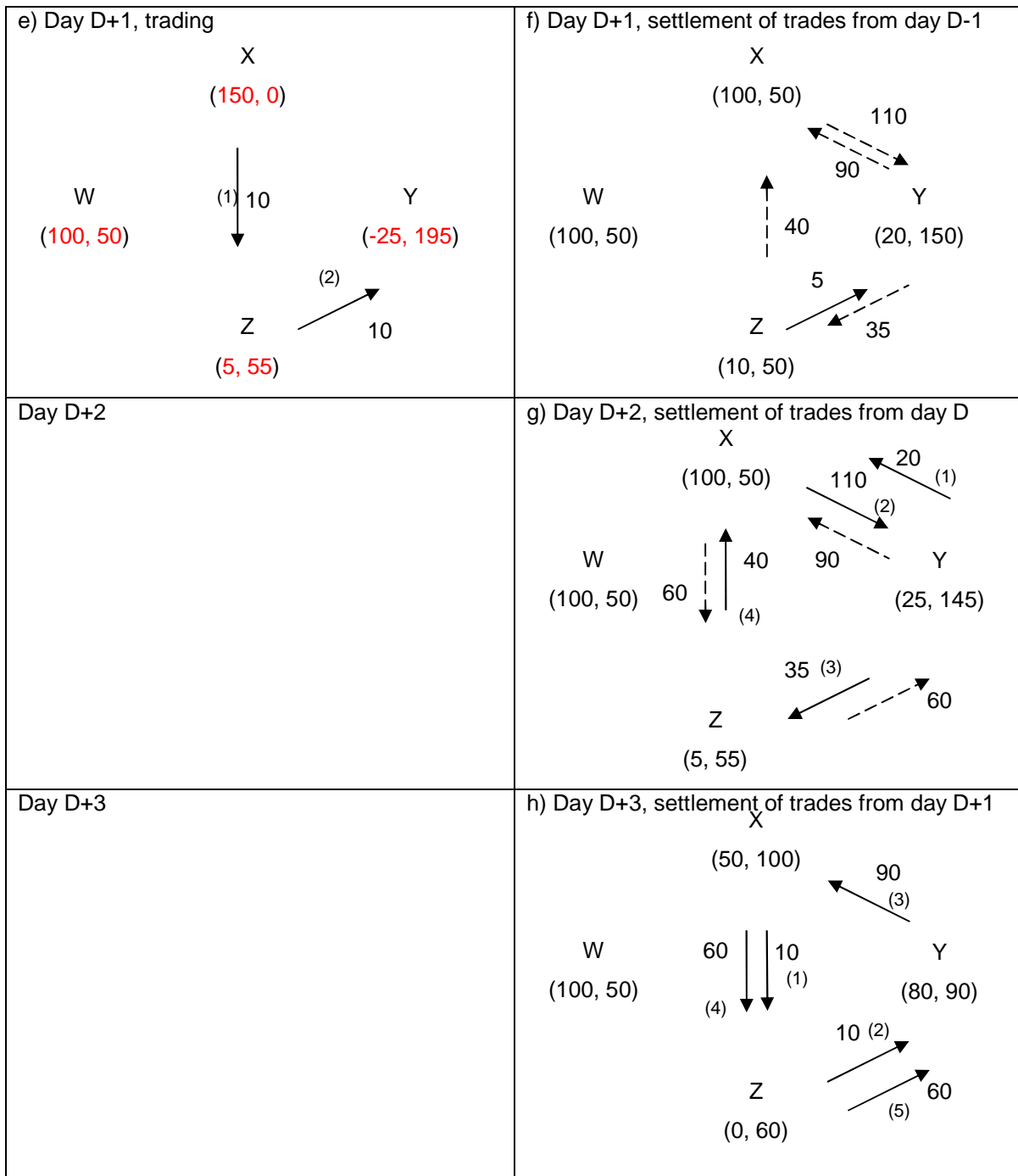
At the start of day D+1, participants use the results from the previous day's settlement (which determined the actual holdings of the security and cash that participants have in their accounts), and they then incorporate into these holdings their expected holdings following the trades from day D-1 and day D, which have not yet settled.⁹ So, X's expected holding of the security is 150 ($100+90-60+20 = 150$) and 0 in cash ($50-90+60-20 = 0$). After trading (panel e), settlement of day D-1 trades (and the trades in the queue) commences (panel f). Only one transaction can be settled (solid arrow). For all other trades participants are either cash or securities-constrained.

To keep the example as simple as possible, we assume no further trading after day D+1; however, we examine the results of settlement of all trades which are still unsettled. Trades from day D will be presented for settlement on day D+2 (panel g). Note that the actual holdings of securities and cash, which are used for settlement, differ from the expected holdings on which the trades from day D were based. For example, at the time of trading on day D, X expected to have (120, 30), while the actual holdings for settlement of these trades are (100, 50). This difference results in the persistence of unsettled trades. Only four trades settle on day D+2; three remain unsettled. After settlement, X ends up with 50 of the security ($100+20-110+40 = 50$).

In this example, it takes until day D+3 before all trades settle. Due to the initial default, settlement efficiency has remained below 100% for three consecutive days. This is caused in part by the trades that are deleted due to the default (the direct effect) and in part by participants trading on the basis of expected holdings of securities and cash which, ex post, turn out to differ from their actual holdings (due to the indirect effects of default). These differences occur whenever some trades unexpectedly do not settle on the anticipated day.

⁹ We assume that participants disregard trades that are already in the queue of unsettled trades. This is due in part to the fact that participants do not know when and if these trades will settle.

Trades (based on participants' expected holdings of cash and securities)	Settlement (given the actual holdings of cash and securities in participants' accounts at time of settlement)
<p>a) Day D-2, trading</p> 	
<p>b) Day D-1, trading</p> 	
<p>c) Day D, trading</p> 	<p>d) Day D, settlement of trades from day D-2</p> 



4 MODEL

As in the above example, we model a SSS with DVP and gross settlement, where settlement occurs with a two-day lag. There are N participants and K securities. All securities prices are assumed to be fixed and normalized to one.¹⁰ The total quantity of each security is also normalized to one. In Section 4.1 we describe the main features of the model and underlying intuition. Section 4.2 details the notation and assumptions.

¹⁰ Allowing for changing securities prices would not affect our qualitative results on settlement failures.

4.1 Description

Initial endowments and timing of events. Participants are randomly allocated initial quantities of cash and securities. We compare two different allocation schemes (described in Section 4.2) with differing degrees of concentration, which allows us to investigate the importance of concentration on the impact of default by the largest participant.

Three "events" occur during each day t in the following order: (i) participants' determination of their expected holdings of cash and securities, which will form the "budget constraints" used for trading on day t ; (ii) trading; and (iii) settlement at the end of day t of trades undertaken on day $t-2$.¹¹

Determination of "expected" budget constraints. As in the stylized example, because settlement occurs with a lag, the budget constraints that are used for trading will represent participants' expected holdings of securities and cash. The expected holdings at the beginning of day t will be the amounts of securities and cash that participants believe will actually be in their accounts following settlement of the trades from days $t-2$ and $t-1$. We make the distinction between participants' expectations in "normal" times and in "crisis" times; i.e., after a participant has defaulted.

Expectations in normal times: We assume that as long as no defaults have occurred, participants expect that all of their previously committed trades will settle (which will actually turn out to be the case). Thus, participants' expected budget constraints at the beginning of day t (reflecting the expected results of settlement of all trades undertaken prior to day t) will be identical to the actual amounts of securities and cash that will be deposited in their accounts once settlement of all trades up to day t has occurred. This means, further, that the amounts of securities and cash actually in participants' accounts on day $t+2$ and used for settlement of day- t trades will be identical to the amounts that were reflected in the expected budget constraints used for determining day- t trades. Thus, no settlement failures will occur.

Expectations in crisis times. As in the example of Section 3, when a participant defaults, all of its unsettled trades are deleted from the system, which leads to the direct effects of the default. As before, there may also be indirect effects of the default. As is discussed in more detail in the next subsection, we assume that participants adjust (reduce) their expected holdings of securities and cash in response to observed settlement failures following the default.

Trading. Trades are assumed to occur randomly. That is, a given security and pair of participants are randomly selected, then the set of feasible trades of the security between the two participants is determined via the two participants' expected budget constraints. Once the set of feasible trades has been determined, a trade is then randomly selected from this set. The expected budget

¹¹ In practice, settlement of day $t-2$ trades will typically begin during the day on day t ; however, completion of the settlement process will generally not occur before the end of the day. For modelling purposes, we assume that settlement of all day $t-2$ trades occurs at the end of day t .

constraints of the participants are then updated to reflect the trade. We compare results from scenarios where trades occur between all possible pairs of participants and securities and where the quantity of trades is limited.

The assumption of random trade behavior is more realistic than might appear to be the case at first glance. Large securities firms are often dealers who trade on behalf of their clients. Trades are executed according to the demands of the clients; therefore, the trades look random from the point of view of the securities firm.

Settlement. As in the example of Section 3, settlement of trades is assumed to occur in the same order as the order in which the trades were undertaken, which maximizes settlement efficiency.¹² A further aid to settlement is the assumption of a queue of unsettled trades, which also reflects practice in SSS's. The settlement process during a given day is assumed to consist of five batches, or iterations. If a trade does not settle in the first batch, it is placed in a queue. Subsequently all other trades are either settled or added to the queue. When all trades have been tried once, the trades in the queue are presented for settlement in the next batch. This process continues until either all trades are settled or all five iterations have taken place.¹³ Trades that are still unsettled at the end of the five batches are placed in the queue for settlement in the next day's settlement process.

Another feature of the model that can reduce settlement failures is the provision of intraday credit, which may be drawn upon during the settlement process. Note that at the point when trades for day t are settled (i.e., at the end of day $t+2$), the holdings of securities and cash that participants have in their accounts reflects the settlement of all trades that were undertaken up to day t . When intraday credit is available, a participant who is short in cash for settlement of a trade can draw on the credit during the settlement process and avoid settlement failure. Simulations with differing amounts of credit availability are compared in Section 5.

As noted earlier, one of the assumptions underlying the model is that there is no securities lending and borrowing program. The difficulty of maintaining a well functioning program is alluded to by Fleming and Garbade (2002). These authors note several potential solutions for helping to resolve a shortage of securities; however each solution has its limits. Among the potential solutions are: i) response by the Treasury to the shortage in the securities lending pool by reopening on-the-run notes. ii) extension of central bank securities lending; and iii) an increase in penalty fees for settlement failures. However, the effectiveness of these solutions is limited by the following considerations: the solutions may only solve settlement fails for particular securities; limits on lending and borrowing may apply; treasuries will need to be involved; and participants may still be unwilling to lend securities in stress periods. In fact, the potential role of a central bank in case of

¹² In practice, SSSs do not actually know the order of trades, although they do know the order in which trades entered into their computer system. They use a number of algorithms to maximise settlement efficiency and to minimise the amount of liquidity that must be provided.

settlement fails may be particularly narrow, especially for the case of SSSs outside the central bank that settle non-treasury securities. Such securities cannot be issued by the treasury and are generally not available in the portfolio of the central bank.

Initial shock. The initial shock in settlement is assumed to stem from an exogenous default of the largest participant, where size is measured by the volume of trades.¹⁴ This does not imply, however, that solvency risk is playing a role in the model. The simulation takes into account liquidity risk only, gauged in terms of the trades that fail to settle because of insufficient cash or securities holdings by the transaction participants. Unlike the interbank contagion literature, participants' losses due to failed trades are not compared with a solvency constraint.

4.2 Notation and assumptions

The following table presents notation used in the model.

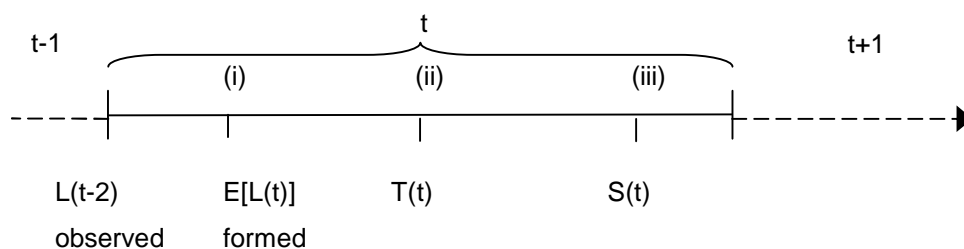
N	number of participants
K	number of securities, with the quantity of each security normalized to 1. If a participant holds 0.01 of security k, it holds 1% of the total outstanding amount of this security.
t	time index, representing one day.
T(t)	Three-dimensional trade flow matrix (N,N,K) of trades occurring during day t, where the entry T_{ijk} represents the quantity of security k that participant i buys from participant j on day t. Negative values equal sales by i to j, and $T_{ijk} = -T_{jik}$. $T_{ijk} = 0$ when no trade has occurred.
S(t)	Three-dimensional settlement flow matrix (N,N,K), whose elements are defined analogously to those of T(t-2), which contains all trades from day t-2 to be settled on day t. ¹⁵ Unsettled trades enter as zeros in the matrix.
L(t-2)	(N, K+1) matrix containing the legal (or post-settlement) holdings of the K securities and cash by each participant at the beginning of day t-2, but <i>only known at the beginning of day t</i> , (or equivalently, after the settlement process at the end of day t-1). The matrix reflects the results of all past settlement; i.e. $L(t-2) = S(t-1) + S(t-2) + S(t-3) + \dots$
$E[L(t) L(t-2), T(t-2), T(t-1)]$	Expectations of L(t) formed by each participant prior to trading on day t and based upon information on all previous settled and unsettled trades. Because L(t-2) will already have been determined by the beginning of day t, $E[L(t)]$ will ultimately differ from L(t) only if some of the participant's trades on days t-2 and t-1 do not settle (which will become known at the end of days t and t+1, respectively).

¹³ Because trades that are not settled in the first batch often settle in a subsequent batch, allowing for a queue of unsettled trades reduces the negative impact on settlement of default by a participant.

¹⁴ In our model, the participant with the largest initial endowment of securities and cash ends up being the participant with the largest amount of trades.

¹⁵ Plus any unsettled trades from day t-3 that are still in the queue.

The figure below depicts the timing of events during a single day t as described in Section 4.1, and the matrices that are relevant for each event.



Expected holdings $E[L(t)]$ are based upon information available at the beginning of day t , which includes the history of trades and settlement, as reflected in $L(t-2)$, $T(t-2)$ and $T(t-1)$. Once $E[L(t)]$ is determined, participants trade, and the trades are recorded in the matrix $T(t)$. After each trade has been determined, the trading participants' cells in the expected budget constraint matrix $E[L(t)]$ are updated before a new trade occurs. At the end of the day, trades that were committed on day $t-2$ are presented for settlement. By comparing $T(t-2)$ with the actual holdings reflected in $L(t-2)$, it is determined if a particular trade is settled or not. Hence $S(t)$ becomes known.

4.2.1 Initial securities and cash holdings, $L(1)$

Securities are initially allocated according to one of two possible allocation schemes: a "diversified" and a "concentrated" scheme. In the diversified scheme each of the N participants receives the quantity $1/N$ of each of the K securities. In the concentrated scheme, the entire quantity of each security is randomly allocated to a given participant. The diversified scheme ensures that participants have equal market shares, while the concentrated scheme leads to differing market shares. Cash positions are assumed to equal a fraction $C < 1$ of the initial endowment of securities.¹⁶ This reflects the idea that cash bears no return and is only held for trading purposes.

Participants can go short in cash during settlement by using credit granted by the CSD. The credit limit is set as a fraction λ of the initial endowment of assets. The credit in the model can be thought of as representing either the collateralised credit provided by a CSD or liquidity that may otherwise be available through the interbank market (which is not formally modelled here). However, since credit is costly, we assume that participants try not to use it on "normal" days. That is, participants do not include credit in their expected budget constraints used to determine their trades. Participants are assumed to make use of their credit line within the SSS only as a backup facility, that is, during the settlement process to avoid unanticipated settlement failures.¹⁷ Note that for day 1, $E[L(1)] = L(1)$ by definition.

¹⁶ Participants who were not endowed with any security receive the quantity C in cash.

4.2.2 Expected budget constraints $EL(t)$

As mentioned in Section 4.1, in normal times participants expect that all trades will settle. This implies that in normal times $E[L(t)] = L(t-2) + T(t-1) + T(t-2)$. In crisis times (i.e., after a participant has defaulted), participants must adjust their expected holdings of securities and cash to reflect the direct and indirect effects of default. Each participant takes account of the direct effects of a default by adjusting $E[L(t)]$ to reflect the fact that all unsettled trades with the defaulting participant will be deleted from the settlement system.

With respect to the indirect effects of default, we assume that on a given day t following a default by a participant, the nondefaulting participants reduce their expected holdings of securities and cash by some amount $\varepsilon_t = \gamma(1 - \theta_{t-1}^*)$, where $0 \leq \gamma$ and θ_{t-1}^* is the measure of indirect settlement efficiency (defined below) on day $t-1$. The coefficient γ captures the degree of conservatism in participants' expectations. Varying this parameter allows us to compare results when participants reduce their expected holdings by very small amounts (γ close to zero) or larger amounts ($\gamma \geq 1$). We report results in Section 5 for scenarios with $\gamma=0$ and with $\gamma=1$.

While our assumption regarding expectations of the indirect effects of a default is admittedly *ad hoc*, it nevertheless provides an element of "endogeneity" in expectations. For example, if, during a crisis period, participants observe that only 80% of trades settled during the settlement process on day $t-1$, they will reduce their expected holdings of securities and cash, used for trading on day t , to 80% of the holdings they would have had if all of their previous trades would settle (assuming that $\gamma=1$).¹⁸

These expectations, while fairly conservative, may still generate additional settlement failures. To the extent that, *ex post*, the percentage of unsettled trades due to the indirect effects of default is higher than what participants had expected, new settlement failures may occur. This explains in part why settlement failures continue beyond the period of the lag in settlement.

4.2.3 Determining trades $T(t)$

As described above, trades are determined by randomly choosing two counterparties and a security and by random selection of a trade from the set of feasible trades of the security between the two participants. The expected budget constraint for participant i is given by the i -th row of the matrix $EL(t)$. The first K columns represent i 's expected quantities of each of the K securities, and the $K+1$ st column represents i 's expected holding of cash. The expected budget constraints of the two participants are used to determine the set of feasible trades of the security. Thus, the maximum amount of security k that participant i can purchase from j is given by:

¹⁸ Allowing participants to draw on their credit lines for trading would not change any of the qualitative results in the model. It would simply widen budget constraints used to determine feasible trades.

$$P = \min[EL(t)_{i,k+1}; EL(t)_{j,k}].$$

The maximum amount of security k that participant i can sell to j is given by:

$$S = \min[EL(t)_{i,k}; EL(t)_{j,k+1}].$$

All feasible trades of security k between participants i and j can then be represented by the interval $[-S,P]$, where, by convention, we assign negative values to sales.

The trade, which will be recorded as entry T_{ijk} in the matrix $T(t)$, is randomly chosen from the interval $[-S,P]$ by use of a symmetrical beta distribution with parameter $\beta > 0$. This distribution has the advantage that varying the parameters of the distribution leads to more or less "extreme" trades; i.e., how close the trade is to the boundaries of the feasible set of trades and, therefore, of the participants' expected budget constraints. As noted in the Appendix, the standard uniform distribution is a special case of the beta distribution, where $\beta = 1$. Values of $0 < \beta < 1$ represent extreme trading behavior, as trades frequently occur near the boundaries of the set of feasible trades. A value of $\beta > 1$ represents participants with less extreme trading behavior.

After T_{ijk} is chosen, the expected cash and security positions in the participants' rows in the matrix $EL(t)$ are updated. Then another security and participant combination is selected.¹⁹

To capture the fact that, in reality, all participants do not trade every security with all potential counterparties, we run some scenarios where the number of trades is limited below the maximum possible number. We limit trades by halting trading when the value of total turnover reaches a proportion μ of the aggregate value of securities and cash holdings in the system. In Section 5 we compare results for scenarios where $\mu = 0.1$ and 0.5 , and where there is no limit (i.e., where trades occur between all possible combinations of counterparties and securities).

4.2.4 Determining $S(t)$

$S(t)$ is the matrix containing all settled trades from day $t-2$, as reflected in the trading matrix $T(t-2)$, plus any unsettled trades from the queue. Trades do not settle if settlement would imply any one of the following:

- A counterparty's cash position becomes overdrawn and exceeds the credit limit granted by the CSD.
- A counterparty becomes short in securities.
- One of the counterparties defaulted prior to settlement

¹⁸ We also discuss in Section 5 some simulations where participants have more "conservative" expectations, whereby they reduce their expected security and cash holdings by even more than the fraction of unsettled trades from the previous day's settlement process.

¹⁹ Our algorithm allows a participant to trade a single security multiple times during a given trading day; however, the participant is not allowed to trade the same security multiple times with the same counterparty.

Trades are settled one-by-one (gross settlement) in the same order as they were traded. Each trade is compared with the holdings of cash and securities represented by the relevant entries in the matrix $L(t-2)$, plus available credit.

For example, trade $T_{ijk}(t-2)$ will be settled on day t if it lies in the interval $[-\bar{S}, \bar{P}]$, where \bar{P} is the maximum amount of security k that i can purchase from j , taking into account i 's credit limit.

$$\bar{P} = \min[L(t-2)_{i,K+1} + \text{credit}_i; L(t-2)_{j,k}]$$

Similarly, \bar{S} is the maximum amount of security k that player i can sell to player j , taking into account j 's credit limit.

$$\bar{S} = \min[L(t-2)_{i,k}; L(t-2)_{j,K+1} + \text{credit}_j]$$

and $L(t-2)_{j,k}$ is the quantity of security k in participant j 's account at the beginning of day t and $L(t-2)_{j,K+1}$ is the amount of cash in j 's account at the beginning of day t (idem for i). Credit_j is the credit of j available from the SSS (idem for i).

Trades that are not settled become zero entries in the $S(t)$ matrix. Trades that do settle are filled in and $L(t-2)$ is updated. As noted above, unsettled trades are placed in a queue, and either settled in a subsequent batch during day t or held in the queue for settlement on day $t+1$.²⁰ At the end of the settlement process on day t , the matrix $L(t-1)$ becomes known.

4.2.5 The initial shock

We assume that due to external factors, the largest participant is not able to fulfil its obligations on day D and subsequent days. We also assume that during day D rumours of the imminent default begin circulating in the market, and participants react by avoiding all trades on day D with the troubled participant. This allows investigation of the impact of a default that is "anticipated" and which will result in less of a shock than a completely unanticipated default. Actual failure of the participant is assumed to occur at the close of the trading period on day D , but before the settlement period begins. Hence, before the settlement process begins on day D , all of the unsettled trades of the failing participant from day $D-2$ and $D-1$ are deleted, and settlement of other trades from $D-2$ proceeds. Such a procedure is in accordance with reality, where a bankruptcy administrator or liquidator may block all unsettled trades in order to protect the interests of creditors.

²⁰ Note that as in reality, partial settlement, i.e.; splitting a large trade into many small ones which are then settled separately, is not allowed. Partial netting of several trades in the same transaction chain is also not allowed. In practice, SSS may use "back-to-back" netting; however, applying netting to a chain longer than two participants is often technically infeasible, since the number of netting possibilities increases exponentially as the chain lengthens.

4.2.6 Calculation of settlement efficiency and market liquidity

We use settlement efficiency as an aggregate measure of liquidity risk. Settlement efficiency is determined by dividing the aggregate value of settled trades by the aggregate value of trades needing to be settled. We distinguish between two measures: total settlement efficiency ($\theta(t)$) and indirect settlement efficiency ($\theta^*(t)$). In the first measure the denominator includes all trades committed two days earlier, including those involving the defaulting participant.²¹ On the other hand, for the measure of indirect settlement efficiency $\theta^*(t)$, the denominator excludes trades involving the defaulting participant. Hence, indirect settlement efficiency is a measure of contagion in the settlement system.

From a financial stability perspective, settlement efficiency is not the only potential measure that is important. Market liquidity matters, too. For example, settlement efficiency may converge very fast to its pre-default levels, while at the same time trading volume is very low because participants limit their trades. Hence, we also construct a second indicator of efficiency, which attempts to measure the fall in market liquidity as a result of the default. Our measure of market liquidity on day t equals the value of turnover on day t as a proportion of the average daily turnover during the ten days prior to the default. Trades from the defaulting participant are excluded from both the numerator and the denominator of this measure.

5 SIMULATION RESULTS

5.1 Parameters

Simulations have been undertaken for the following combinations of parameter values:

- Number of participants: $N = 5$ or 15 ;
- Number of securities: $K = 20$ or 30 ;
- Values of the Beta distribution: $\beta = 0.2$ or 1 ;
- Multiplier (of prior indirect settlement inefficiency) for the adjustment of expected holdings of securities after default): $\gamma = 0$ or 1 ;
- Availability of credit (as a fraction of value of initial endowments): $\lambda = 0, 0.5$ or 1 ;
- Amount of cash held, as a proportion of total value of initial endowment of securities: $C = 0.05$;
- Limit on total turnover as a proportion of aggregate value of securities and cash: $\mu = 0.1, 0.5$, or no limit;
- Allocation of initial endowments: “diversified” and “concentrated” scheme.

One thousand simulations have been run for each combination of parameter values. Each simulation runs for 20 days prior to a default occurring. This ensures that trading behavior has

²¹ Plus all the trades that did not settle previously; i.e., the ones in the queue.

“stabilized” well before default occurs. At the end of the 20th day, the largest participants default. The simulation continues for ten days following default.

Although N and K have been set at relatively low levels, these values are in fact more realistic than may first appear. Whereas the number of securities traded in actual SSSs may be very large, often only a small number of securities is actively traded. Similarly, even when an SSS has many participants, it is common that only a few active participants, such as large custodians, broker/dealers, central counterparties or specialized traders, account for a majority of the trading. In addition, many SSSs actually involve only a dozen participants.

5.2 Results

5.2.1 Market shares

Table 1 reports the market share on day D-2 of the largest participant (which will be the defaulting participant) for differing values of N and initial allocation schemes, and for a scenario with the following values of other parameters: $K=30$; $\mu = 0.5$; $\beta = 1$. Market shares are calculated as the (absolute value of) total trades of the participant as a proportion of the (absolute value of) total trades in the system on a “normal” trading day. As expected, the average market share of the largest participant is higher for lower values of N and for the concentrated allocation scheme. Interestingly, varying the values of K , β and μ does not significantly affect the market shares of the largest participant.²² Market share appears to be determined by the concentration of initial endowments, which is largely determined by N and the initial allocation scheme. The reported market shares are in line with what is observed in many SSSs in Europe, where often a few participants generate the largest proportion of the business.

	N=5		N=15	
Initial endowment allocation scheme	Diversified	Concentrated	Diversified	Concentrated
Average	0.20	0.25	0.07	0.11
Std.dev	0.02	0.03	0.01	0.02
Max	0.31	0.35	0.12	0.20
Min	0.11	0.15	0.03	0.04

Note: results reported for $\mu = 0.5$, $\beta = 1$, $K = 30$.

5.2.2 First day impact (D-Day)

The largest participant is assumed to default on all of its outstanding obligations on day D. Tables 2 and 3 illustrate the impact on settlement efficiency (total settlement efficiency θ and indirect settlement efficiency $\hat{\theta}$) for differing values of N , β , μ and λ and for different initial allocation

schemes.²³ Several observations can be made. First, θ and θ^* vary considerably across different parameter combinations. The value of average θ (total settlement efficiency) ranges from 32% to 83%, while θ^* ranges between 64% and 96%. Settlement efficiency also varies considerably across simulations for a given combination of parameter values, as can be seen from the relatively high standard deviations.

Table 2: Average value of settlement efficiency on day D as a function of credit limit and initial endowments (standard deviations between brackets)

Initial endowment allocation	Credit limit (λ)	N=5		N=15	
		Total settlement efficiency (θ)	Indirect settlement efficiency (θ^*)	Total settlement efficiency (θ)	Indirect settlement efficiency (θ^*)
Concentrated	0	32.26 (10.82)	64.33 (17.59)	57.84 (8.99)	75.69 (9.58)
	0.5	43.49 (8.23)	86.71 (6.84)	72.27 (5.62)	94.89 (2.58)
	1	43.22 (8.21)	87.02 (6.92)	72.19 (5.95)	94.926 (2.53)
Diversified	0	42.12 (10.58)	70.16 (15.14)	70.20 (7.34)	80.95 (7.48)
	0.5	53.96 (7.28)	89.87 (5.56)	83.95 (3.80)	96.99 (1.74)
	1	53.80 (7.44)	89.71 (5.68)	83.99 (3.72)	96.97 (1.68)

Results reported for $K=30$, $\mu = 0.5$, $\beta = 0.2$.

As expected, total settlement efficiency is low when the market share of the largest participant is high (N=5 and concentrated allocation). Settlement efficiency (both direct and indirect) is positively related to N, the number of participants. This can be explained by the fact that the higher is N, the smaller the initial shock, as the market share of the largest participant is lower. In addition, higher values of N imply that the shock will be distributed among more participants, thereby raising indirect settlement efficiency.

Liquidity, in the form of credit, also appears to be an important tool in limiting contagion (a higher λ corresponds to a higher credit limit). Importantly, however, whereas increasing λ from 0 to 0.5 has a significant effect on settlement efficiency, further increasing λ from 0.5 to 1 appears to have

²² For very small values of K, increasing the number of securities does have a significant impact, causing concentration to decline, all else equal. However, the magnitude of the impact decreases as the initial number of securities increases, so that increasing the number of securities beyond 30 has only a negligible effect.

²³ Note that it is not useful to report the other liquidity measure, market liquidity, for the first day of the crisis. On day D, participants only stop trading with the defaulter. Hence, total trade volume is diminished by the market share of the largest participant. Participants only restrict their trades with the other counterparties on subsequent days, after observing $\theta < 100$.

no additional impact. This is possibly because even generous liquidity provision cannot completely eliminate settlement failures for participants who find themselves short in securities.

At first sight, the level of contagion – as measured by θ^* – may appear to be very limited once credit is available. Indeed, in cases where $\lambda \geq 0.5$, θ^* is always above 85% and even reaches 96% with the diversified initial allocation scheme. However, these are average values across simulations, and they reflect the first day on which contagion effects may occur. As will become clear below, the level of disruption can be much higher in subsequent days. Also, given the sizeable transaction volume in securities markets, these figures are non-negligible. For example, many SSSs settle daily transaction volumes of more than 100bn EUR, which implies that a drop of 10% in settlement efficiency represents 10bn EUR in trades.

Table 3 reveals that the degree of contagion on the first day of default can be significantly worse if average turnover prior to default (μ) is higher or if trading behavior is more extreme (lower β). Settlement efficiency (both θ and θ^*) is higher for higher β (less extreme trades) and lower μ . When participants trade more at the boundaries of their budget constraints, the greater is the likelihood that they will experience an unanticipated short position in cash or securities during a crisis; hence, the likelihood of a settlement failure is increased. In addition, settlement efficiency decreases when turnover increases (higher μ). This is due to the fact that the chain of trades becomes longer when turnover increases (i.e. the number of times a single security is traded during a given trading day increases). This causes the contagion effects of a single settlement failure to increase. This result suggests that more liquid markets may actually result in more significant contagion.

Table 3: Total settlement efficiency on day D as a function of turnover and trade behavior

Extremeness of trades (β)	Credit limit (λ)	$\mu = 0.1$		$\mu = 0.5$		Turnover not limited	
		Total settlement efficiency (θ)	Indirect settlement efficiency (θ^*)	Total settlement efficiency (θ)	Indirect settlement efficiency (θ^*)	Total settlement efficiency (θ)	Indirect settlement efficiency (θ^*)
0.2	0	64.24	84.70	57.84	75.69	43.75	57.44
	1	74.77	98.47	72.19	94.92	62.67	82.35
1	0	71.09	92.43	68.10	88.05	60.79	78.64
	1	76.12	98.92	74.69	96.66	70.03	90.61

Results reported for N=15, K=30, concentrated allocation. Qualitative results are similar for diversified allocation and other values of N and K.

5.2.3 Net buy vs. net sell position of the defaulter

Given the need for cash in every transaction, we expect there to be a relation between the net trade position of the defaulter and settlement efficiency. The default of a net buyer extracts cash from the system. As cash is used in every transaction, this may lead to more significant contagion, and hence lower settlement efficiency, than when the defaulter is a net seller.

Suppose that participant i is the defaulter. The measure for the net trade position of participant i controls for the volume traded and is defined by:

$$\frac{\sum_{j \neq i}^N \sum_{k=1}^K T_{ijk}}{\sum_{j \neq i}^N \sum_{k=1}^K |T_{ijk}|}$$

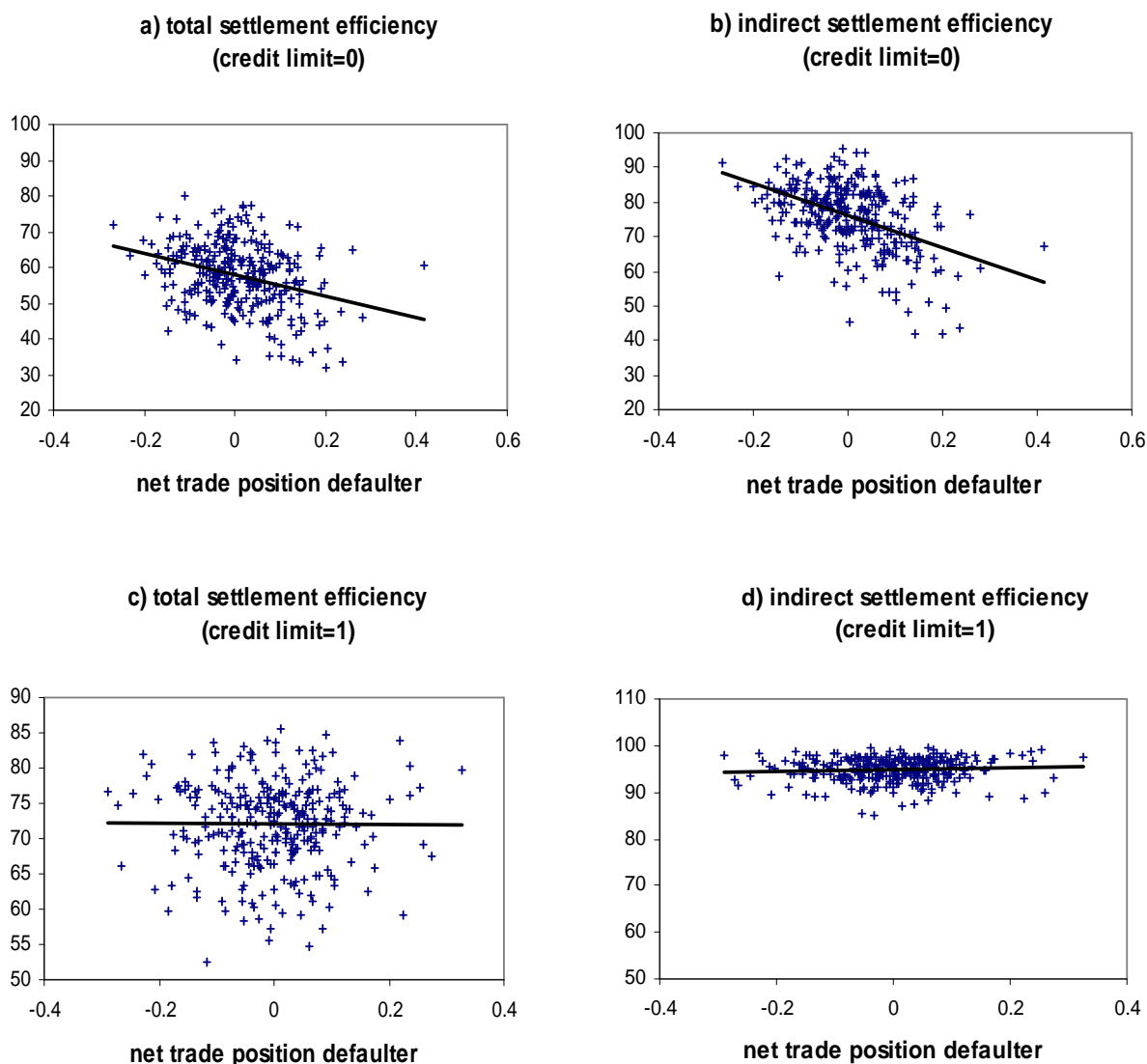
Note that a negative value of T_{ijk} represents a sale of security k by participant i to j . Similarly, a positive value denotes a purchase of security k by participant i from j . Thus, a negative value for the trade position signifies that participant i is a net seller and a positive value signifies that i is a net buyer. For example, a value of 0.1 implies that participant i bought 10% more securities than it sold and hence is accumulating securities in exchange for cash (net buyer).

Figure 1 (a, b, c, d) plots total and indirect settlement efficiency against the net trade position of the defaulter on trades from day D-2. All of the graphs in this figure are generated from a scenario where $N=15$, $K=30$, $\mu = 0.5$ and $\beta = 0.2$ and the concentrated initial allocation scheme.²⁴ Figures 1a and 1b illustrate the case where no credit is available ($\lambda = 0$). These figures illustrate that the net trade position of the defaulter has an important impact on settlement efficiency. A net buy position of the defaulter has a more significant negative impact on settlement efficiency than a net sell position, and in some cases the net buy position leads to a near complete breakdown of settlement. As might be expected, the effect of the net buy position is stronger for the measure of indirect than total settlement efficiency.

Figures 1c and 1d illustrate that once participants are able to draw on credit lines during settlement (higher λ), the impact of net trade position on settlement efficiency disappears. Importantly, however, settlement failures still occur, despite generous liquidity provision. As already noted, this is due to the fact that securities transactions contain a securities leg as well as a cash leg. Due to the initial default, some securities remain with the defaulting participant, causing some participants to be short in securities. Even unlimited credit cannot make up for these short positions.

²⁴ Qualitative results are similar for other values of β, μ, N, K and the diversified allocation scheme.

Figure 1 (a, b, c, d): net trade position and settlement efficiency



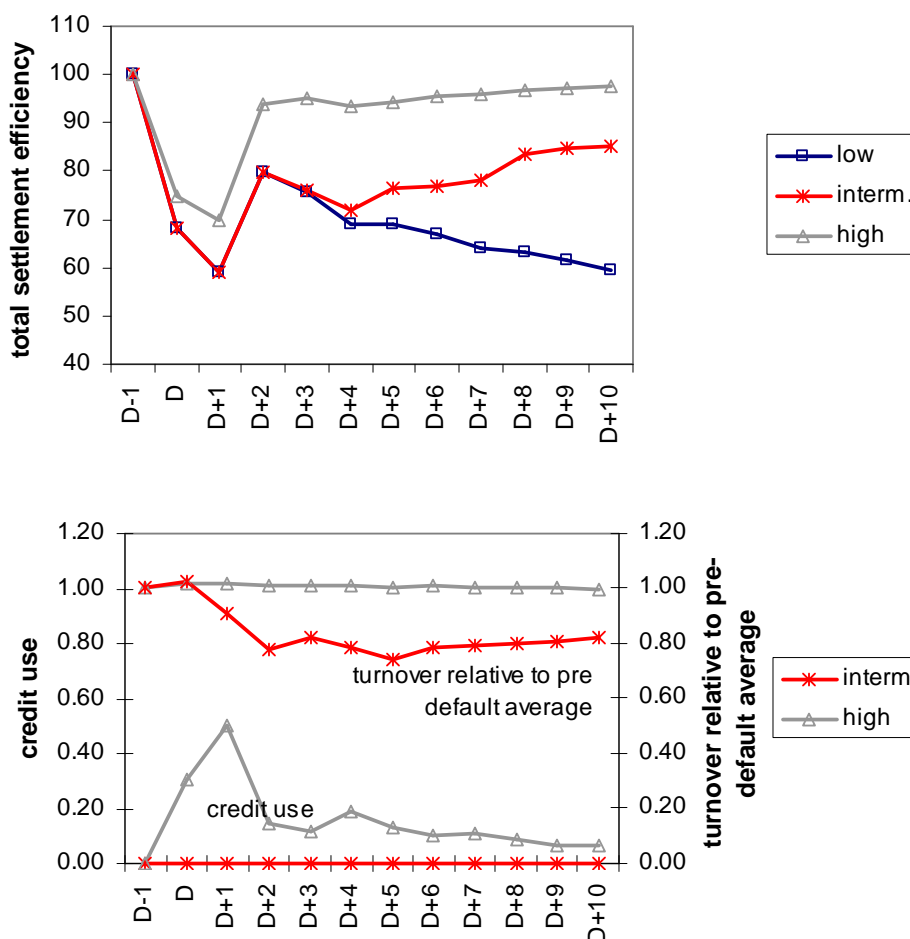
Note: $N=15$, $K=30$, $\mu = 0.5$ and $\beta = 0.2$. For illustrative purposes, only 300 random observations out of the 1000 simulations are plotted.

5.2.4 Length of the crisis

We have argued above that while the impact of a disruption on the first day is important, it is also of interest to know what happens in subsequent days. This is especially true if settlement efficiency is used as a measure of liquidity risk in SSSs. For example, if a large proportion of the unsettled trades resulting from the first day of a shock can be settled the next day, then replacement cost risk and liquidity risk will be judged to be limited, as settlement will have only been delayed by one day. However, if settlement failures persist, uncertainty will remain and liquidity risk or replacement cost risk may become significant.

Figure 2, which contrasts three scenarios, provides information regarding the length of the crisis. The "Low settlement" scenario is one where no liquidity is available ($\lambda = 0$) and where participants make no adjustments to their expected asset holdings to account for the indirect effects of default ($\gamma = 0$). The "Intermediate settlement" scenario is one where liquidity is still unavailable but where participants adjust their expected asset holdings to account for the indirect effects of default ($\gamma = 1$). The "High settlement" scenario represents one where liquidity is abundant ($\lambda = 1$) but where participants do not adjust their expected asset holdings. (Results for this scenario are not significantly different if participants do adjust their expected holdings.) Values of other parameters that are constant across all three scenarios are: $N=15$, $K=30$, $\mu = 0.5$, $\beta = 1$, and concentrated initial allocation scheme.

Figure 2: Length of the crisis



The top panel of Figure 2 presents settlement efficiency over a period of several days (from D-1 up to D+10) for each of the three scenarios. Only total settlement efficiency is plotted, since from day D+2 onwards there are no more trades involving the defaulted participant, and indirect settlement efficiency becomes equivalent to total settlement efficiency. The lower panel of Figure 2 presents the volume of credit extended, as well as the level of market liquidity. Results in this panel are presented only for the intermediate and high scenarios, as no credit is available in the low scenario,

and there is also no fall in market liquidity, since participants do not adjust their expected holdings of assets following default.

The upper panel of Figure 2 clearly demonstrates that even in a SSS with DVP and gross settlement, there is still a possibility of a significant, multi-period disruption of settlement activity when a large participant fails. This is the case even when the default is anticipated by the market (as is the case in our simulations). In the Low scenario, settlement efficiency falls drastically and does not improve throughout the ten-day period following default. This is due to the impact of the initial settlement failures and the absence of both liquidity and adjustment of expected asset quantities. The more trades that fail to settle in this case, the greater the discrepancy between participants' actual holdings of securities and cash at the point of settlement and the expected budget constraints they used for determining the trades. Hence, the more trades that fail to settle, the more likely it is that future trades will fail to settle (and settlement efficiency will fall), unless the situation can be corrected for by liquidity provision.

This is in sharp contrast with the High scenario, where the drop in settlement efficiency is smaller at the beginning of the crisis and the increase in efficiency is more dramatic. The rapid increase in settlement efficiency in this scenario occurs as a result of ample liquidity provision by the SSS. In the Intermediate scenario, settlement efficiency is identical to that for the Low scenario up to day D+3, after which settlement efficiency improves as a result of participants' downward adjustment of their expected budget constraints (in response to observed settlement inefficiency from the previous day). This scenario illustrates that even in the absence of liquidity provision, settlement efficiency can improve due to participants limiting their expected budget constraints and, consequently, the volume of their trades.

In each of the three scenarios, settlement efficiency is lower on the day after the initial default than on the day itself. This is because on day D+1 there is not only the direct impact of the default but also the subsequent indirect effect on the day D-1 trades. On day D+2 settlement efficiency improves relative to D+1, as there are no longer any unsettled trades involving the defaulting participant; hence there are no longer any direct effects of the default.

The Intermediate scenario suggests that participants' adjustments of their expected asset holdings and the resulting limitation of the volume of trades can serve as a partial substitute for liquidity provision in raising settlement efficiency. This raises the question as to whether more conservative expectations alone (e.g., $\gamma \gg 1$) could lead to settlement efficiency as high as that occurring in the High scenario. This turns out not to be the case. When expectations are very conservative, trading virtually halts. As a result, there are no new trades which, when settled, would then allow the trades in the queue to settle. Hence, unsettled trades remain in the queue while trading volume is very low, resulting in very low settlement efficiency.

An important observation arising from all three scenarios illustrated in the upper panel of Figure 2 is that even with generous liquidity provision, settlement efficiency does not return to its "pre-stress

event" levels within 2 days of the crisis. Even if participants lower their expected holdings of assets to account for contagion effects (as in the Intermediate scenario), it is still possible that, *ex post*, the actual holdings of cash and securities after settlement will be lower than assumed, and that further settlement failures will occur. This is indeed what happens in the simulations, especially when no credit is available ($\lambda=0$) and, by definition, when participants make no adjustments to their expected asset holdings (the Low scenario).

Restoring settlement efficiency through either credit provisioning or conservative trading volume comes at a cost. This is illustrated in the lower panel of Figure 2, which presents the use of credit and the evolution of market liquidity. The latter measure reveals that in the Intermediate scenario (the only scenario in Figure 2 where participants adjust their expected budget constraints) average trade volume falls to around 80% of the pre-default level on day D+1. In subsequent days, market liquidity improves, although only slightly. In the High scenario, aggregate end of day credit usage – as a percentage of outstanding securities – peaks on day D+1 at around 50% of the aggregate value of asset holdings, then diminishes as settlement efficiency restores. It nevertheless remains at a level well above 10% in the subsequent days.

The fact that liquidity provision and participants' expectations are partial substitutes in restoring settlement efficiency suggests a trade-off from a financial stability perspective. Generous liquidity provision places a heavy burden on the liquidity provider but does not reduce trading activity, while conservative reactions by market participants avoid the burden on the liquidity provider but entail a fall in trading activity (resulting in less liquid markets). From a financial stability perspective a balance will need to be struck between the two. Also, the timing of the impact of the measures is different. Liquidity provision increases settlement efficiency immediately, while participants' reactions affect settlement efficiency only with a two-day lag.

The above discussion has suggested that liquidity provision by the SSS (or a central bank supporting the SSS) is an important policy tool for improving settlement efficiency in periods of market disruption. Access to liquidity in times of stress loosens participants' cash constraints, resulting in higher settlement efficiency. Unfortunately, however, this solution may not always be possible, as generous liquidity provision may be judged by the SSS to be too costly or too risky. Even if the SSS is supported by a central bank, there may be limits on the amount of credit that the central bank is willing to provide, as such an amount might have an impact on monetary policy objectives.

6 CONCLUSIONS, POLICY IMPLICATIONS AND ONGOING RESEARCH

This paper has demonstrated that liquidity risk may be important in SSSs with gross settlement, even in systems with delivery versus payment and generous liquidity provision. Although DVP systems eliminate principal risk in SSSs, DVP does not eliminate replacement cost and liquidity risk. Moreover, settlement disruptions may persist over a period of several days. The analysis uses a multi-period model to analyze the extent and dynamics of settlement failures that may occur as a

result of the default of the largest participant. From a financial stability point of view, it is important to understand the mechanics of breakdowns in settlement efficiency, the factors exacerbating disruptions, and the policy tools that may help to resolve crises.

The results suggest that settlement failures due to the default of a large participant are higher in SSSs with a limited number of participants and a relatively high volume of trading. The trading behavior of participants also appears to be important. The more "extreme" are trades – in the sense of being close to the boundaries of participants' expected holdings of cash or securities – the greater the degree of settlement inefficiency induced by a default. Extremeness of trading behavior can also be linked to the types of participants observed in practice. Participants with less extreme trades may be thought of as those trading for their own accounts, mainly initiating a limited number of transactions in buy and hold positions. Participants with more extreme trade behavior may be thought of as broker/dealers, initiating many trades on behalf of their customers. In order to limit costs, these participants trade at the limits of their budget constraints, holding relatively small amounts of surplus cash and securities in their own books.

The trading behavior of the defaulting participant can also have an important impact. A defaulter who has a net buy position will cause more settlement failures than a defaulter with a net sell position, at least when little credit is available. The importance of the net trade position can be explained by the fact that cash is used in every transaction, while a security is only used for transactions involving that security. When enough credit is provided by the SSS, the relative impact of a defaulting institution's net trade position on settlement inefficiency disappears. In other words, liquidity provision by the SSS can help participants to absorb the shock created by the default of a participant, thereby reducing the negative impact of a net buy position of the defaulter relative to a net sell position. This suggests that liquidity provision can be an important policy tool for central banks in supporting the functioning of financial markets. However, injecting enough liquidity in the system to prevent severe contagion of settlement failures may prove to be quite costly or may interfere with other policy objectives. Moreover, because securities transactions involve both a securities and a cash leg, liquidity provision cannot completely eliminate settlement failures due to major market disruptions.

Voluntary limitations on the volume of trades by participants may act as a partial substitute for liquidity provision in alleviating settlement inefficiency due to a market disruption. That is, either fairly conservative reactions by participants to the crisis or ample liquidity can improve settlement efficiency. However, these two alternatives may lead to a trade-off from a financial stability perspective. Liquidity provision places a heavy burden on the liquidity provider but does not reduce trading activity, whereas conservative reactions by market participants avoid the burden on the liquidity provider but may result in less liquid markets.

One type of risk that the model of this paper does not consider and that is potentially important is a fall in securities prices coming about if participants increase their sales of securities in an attempt to

raise cash for (future) settlement of trades that did not settle due to the disruption.²⁵ As noted earlier, such a fall in asset prices would translate into solvency risk for participants rather than feeding back directly into the liquidity risk. Hence, in order to analyze the market risk arising from a disruption in a SSS, it would be necessary to have a model with capital and solvency constraints for participants, in addition to the "budget" constraints used for trading in the model of this paper. This goes beyond the scope of the paper, given the complications that would be involved in conducting our simulations with such an "enhanced" model and given our focus on liquidity risk. Hence, we are unable to analyze policy questions related to the potential effects of market disruptions in SSSs on securities prices and any resulting weakness in participants' balance sheets.

It might also be desirable in future work to allow for securities lending and borrowing programs. It would be necessary, however, to ensure that the size of the securities lending pool is endogenously determined, and to allow for changes in participants' willingness to lend securities during crises.

The result that settlement failures can be severe over a period of days is potentially important information for SSSs. This result suggests that assessments of liquidity risk that only focus on the initial day of the disruption may significantly underestimate the total amount of settlement failures and the ultimate amount of liquidity needed to guarantee timely settlement in case the largest participant fails. One way to shorten the potential length of crises is to try to limit the lag between trade and settlement. If technology could allow for real-time settlement, for example, participants would not need to form expectations about their cash and security holdings. Although settlement failures in response to a major disruption would still occur, multi-day contagion effects would no longer arise.

²⁵ Recall that a sale of a security would increase the cash available to the participant only in two-days' time.

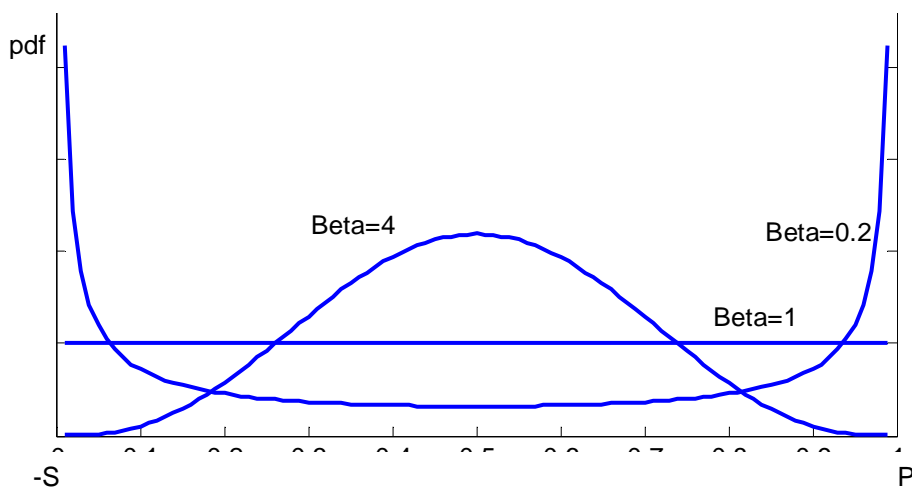
7 APPENDIX

The beta distribution

The beta distribution describes a family of curves that are unique in that they are nonzero only on the interval $[-S, P]$. The shape of the beta distribution is quite variable depending on the values of the parameter β , as illustrated by the plot below. The constant pdf (the flat line) shows that the standard uniform distribution is a special case of the beta distribution. $0 < \beta < 1$ represents participants with extreme trade behaviour as they frequently use the limits of the budget constraint. $\beta > 1$ represents participants with less extreme trade behaviour.

In the simulation, $\beta < 1$, resembling broker/dealers which only hold securities and cash on their account for trading purpose. In order to minimise the costs for their clients, they try to use all their margin.

Figure 3: PDF of the beta distribution



REFERENCES

- Allen F. and Gale D., 2000, "Financial contagion", *Journal of political economy*, vol 108, no 1, 1-33
- Angelini P., Maresca G., Russo D., 1996, "Systemic risk in the netting system", *journal of banking and finance*, 20, 853-868
- Angilini P., 1998, "An analysis of competitive externalities in gross settlement systems", *Journal of banking and finance*, 22, 1-18
- BIS, 1992, "Delivery versus payment in security settlement systems", CPSS
- BIS, 2001, "Assessment methodology for recommendations for Securities Settlement Systems," Report of the CPSS-IOSCO Joint Task Force on Securities Settlement Systems, CPSS publication no. 46.
- Cifuentes R., Ferruci G. and Shin H., 2004, "Liquidity risk and contagion", *forthcoming in Journal of the European economic association*
- De Bandt O. and Hartmann P., 2000, "Systemic risk: a survey", *CEPR discussion paper 2634*
- Degryse H. and Nguyen G., 2004, "Interbank exposures: An empirical examination of systemic risk in the Belgian banking system", *National bank of Belgium working paper*, no 43
- Diamond D. and Dybvig P., 1983, "Bank runs deposit insurance and liquidity", *Journal of political economy*, 91, 401-419
- Diamond D. and Rajan R. 2003, "Liquidity shortage and banking crises", *NBER Working Paper No. w10071*
- Fleming M. and Garbade K., 2002, "When the back office moved to the front burner: Settlement fails in the treasury market after 9/11", *FRBNY policy review*, November
- Furfine, Craig H., 2003, "Interbank Exposures: Quantifying the Risk of Contagion." *Journal of Money, Credit and Banking* 35, February, 111-28.
- Humphrey D., 1986, "Payments finality and risks of settlement failure", in *Technology and the regulation of financial markets: Securities, futures and banking*, edited by A. Saunders and L. White, Lexington, MA: Lexington books, 97-120
- Iori G., 2004, "An analysis of systemic risk in alternative securities settlement architectures", *ECB working paper*, no 404, November
- Kahn C. and Roberds W., 1998, "Payment system settlement and bank incentives", *The review of financial studies*, vol 11, no 4, 845-870
- Kahn C., McAndrews J. and Roberds W., 2003, "Settlement risk under gross and net settlement", *Journal of money credit and banking*, vol 35, no 4
- Leinonen H. and Soramaki K., 1999, "Optimising liquidity usage and settlement speed in payment systems", *Bank of Finland discussion papers*, 16/99, 1-39
- Northcott, 2002, "Estimating settlement risk and the potential for contagion in Canada's automated clearing settlement system", *Bank of Canada working paper 2002-41*
- Rochet J. and Tirole J., 1996, "Interbank lending and systemic risk", *Journal of money credit and banking*, 28, no 4, 733-762
- Upper C. and Worms A., 2004, "Estimating bilateral exposures in the German interbank market: Is there danger of contagion?", *European Econ. Rev.*, Vol. 48, No. 4 (Aug.): 827-849.