BANK OF ENGLAND

Working Paper No. 404
The impact of payment splitting on liquidity requirements in RTGS

Edward Denbee and Ben Norman

October 2010

Working Paper No. 404
The impact of payment splitting on liquidity requirements in RTGS

Edward Denbee[1] and Ben Norman[2]

Abstract

This paper examines the impact that payment splitting could have upon the liquidity requirements and efficiency of a large-value payment system, such as the United Kingdom's CHAPS. Using the Bank of Finland Payment and Settlement Simulator and real UK payments data we find that payment splitting could reduce the liquidity required to settle payments. The reduction in required liquidity would increase as the payment splitting threshold decreased but the relationship is non-linear. Liquidity savings are not homogeneously distributed, with some banks benefiting more than others.

Key words: Payment systems, simulations, payment splitting, liquidity-saving mechanisms..

JEL classification: E42, C63.

Contents

Summary

This paper examines the impact that payment splitting could have upon the liquidity requirements and efficiency of a large-value payment system, such as the United Kingdom's CHAPS. Under payment splitting, a threshold value for payments is defined. Any payments larger than this threshold are split into equal pieces, each smaller than the threshold, and are then settled. In this study we use real UK payments data and the Bank of Finland Payment and Settlement Simulator to test two hypotheses: that (i) payment splitting can reduce the length and impact of payment queues prior to settlement; and, equivalently, (ii) payment splitting can reduce the liquidity requirements of the system.

A number of systems worldwide already adopt payment splitting, either as a formal mechanism or through informal guidance and practice, as a means of being more liquidity efficient. In CLS, a foreign exchange cash settlement system, a currency threshold is set for each currency that it processes. Any eligible transaction above this threshold, in either currency, is split into smaller, equally sized transactions. The Swiss SIC payment system, the Japanese large-value payment system, BoJ-Net and the Canadian securities settlement system, CDSX, all have guidelines or rules that encourage participants to split the largest payments into smaller pieces to aid payment co-ordination and liquidity efficiency.

Our results suggest that if banks were liquidity constrained and, hence, payments were queued prior to settlement, payment splitting could significantly reduce the length of these queues. Splitting allows partial settlement of payments where otherwise none would have been possible. This directly reduces the value of payments queued. Beyond this the recipient bank may be able to use this liquidity to settle queued payments of its own resulting in a favourable 'payment cascade' effect. Reducing the splitting threshold generally results in greater reductions in payment queues.

We also find that, equivalently, payment splitting can reduce banks' liquidity requirements. Splitting payments into smaller pieces and releasing them piecemeal can help banks to co-ordinate their incoming and outgoing payments resulting in less demand for liquidity. By spreading the largest-value payments over time, banks are able to use incoming payments to fund the remaining pieces of an outgoing split payment.

Given the potential benefits from payment splitting, it is worth asking why it is not more widespread. We identify two issues that may discourage systems from adopting payment splitting. First, the liquidity savings that result from this approach are not uniformly distributed. In our simulations, most banks made savings, whereas a few saw an increase in their liquidity needs. The latter tended to be those banks whose payment flows are most dependent on the arrival of incoming payments. In practice we expect that these banks would change their behaviour following the introduction of payment splitting.

Second, we recognise that some legal questions could be raised by payment splitting: above all, if a bank goes into administration after having only partially completed a payment, what is the status of that payment? Whether, and (if so) to what degree, this introduces risk depends upon the type of transaction (if any) that is underlying the payment. While a risk for some underlying transaction types, we conclude that in some cases splitting may actually reduce credit risk. We do not attempt to address the legal questions in detail but merely highlight the issues that a system operator would need to consider if it were to implement payment splitting functionality.

This paper does not seek to propose the adoption of payment splitting functionality in the United Kingdom but rather contributes to the growing literature on mechanisms for making RTGS payment systems more liquidity efficient.

1   Introduction

CHAPS is the United Kingdom's large-value payment system. On average, more than 130,000 payments a day are settled in CHAPS, totalling approximately £280 billion,[1] equivalent to almost one fifth of UK GDP. The system comprises fourteen settlement banks (hereafter 'banks'), which settle payments made to each other in central bank money.[2] It is a real-time gross settlement (RTGS) system meaning that payments are settled with finality, on an individual basis, as they arrive at the RTGS processor at the Bank of England, provided there are sufficient funds on the account. This is in contrast to deferred net settlement (DNS) systems whereby all payments are settled at the end of the day with only net payment values flowing between banks. While RTGS systems eliminate interbank credit risk, as no bank is ever exposed to any counterparty within the payment system, they are costly in terms of liquidity requirements. In a DNS system the liquidity required to settle all payments is a bank's end of day net position if it is a net sender, otherwise the liquidity requirement is zero. The timing of payments received and sent intraday is immaterial. In an RTGS system, however, the variation in the timing of payments sent and received intraday can create significant net sender positions and thus significant liquidity requirements.

There are many ways that this liquidity requirement can be reduced. Recent designs in large-value payment systems, such as the euro-area's TARGET2, BoJ-Net in Japan and BOK-Wire+ in South Korea, have sought to reduce liquidity requirements via the introduction of offsetting algorithms which match incoming payments with outgoing payments and settle them simultaneously.[3] Such settlement requires the net amount of liquidity to settle matched payments rather than the gross amount of the largest payment. There are several other methods that can be employed to save liquidity and limit operational risk in payment systems. Many of these focus on getting payments to be submitted earlier in the day, thereby increasing the time for liquidity recycling and payment offsetting. One such method is throughput guidelines, which are used in CHAPS and CHIPS (in the United States) among others. This paper investigates a further mechanism for liquidity saving: payment splitting, which is a method of saving liquidity applied in some systems, notably the multi-currency cash settlement system CLS.

We simulate payment splitting in the absence of any other liquidity-saving devices. In practice it is unlikely that payment splitting would be introduced as a stand-alone liquidity-saving

---

[1] Bank of England (2009).
[2] Figures quoted in this introduction are based on 2009 H1 data.
[3] For a discussion of changes to large-value payment systems see Bank for International Settlements (2005).

mechanism (LSM). It is also likely that the benefits of payment splitting would complement other LSMs, especially offsetting algorithms. The largest payments may be difficult to offset as they require comparatively large incoming payments to do so (although some countries' payment systems, eg Canada's LVTS, have so-called 'jumbo' offset algorithms for just such a purpose). If they could be split into smaller pieces it may be easier to find offsetting payments and the offsetting algorithm would work more efficiently.

1.1 Payment splitting

There are two ways that payments can be split:

    i)   Payments above a certain size threshold are split into a number of payments each smaller than the threshold; and

    ii)   Payments are split so as to use up all the liquidity available.

The second approach is equivalent to splitting the payment into single units of currency and then settling as much of the payment as liquidity availability allows. When more incoming liquidity arrives, this can then be used to fund the settlement of more units of the split payment. This maximises the amount of liquidity that is released and thus is more efficient than the first method.[4] This method would be optimal from a liquidity efficiency point of view. However, in reality, this is unlikely to be a practical option. There are a number of considerations. Firstly, while not insurmountable, splitting payments into single units would be computationally more complex – a single £1 billion payment split into 1 billion payments of £1 would create much more traffic than CHAPS processes in an average year; secondly, in the event of operational difficulties the reconciliation of potentially millions of payment pieces could be problematic. It has been noted by some banks that concerns about the reconciliation of split payments limit the willingness of liquidity managers to split payments themselves. Besides this, it may not be optimal for a bank to split multiple payments as each submission attracts a settlement fee. Splitting a large payment into numerous smaller ones could be costly. Possibly because of these issues, where payment splitting occurs in practice, the first approach tends to be adopted. We too opt for such an approach in the analysis presented below, splitting payments into equal-sized pieces, each smaller than the threshold.

---

[4] To take an extreme example, if two identical payments were each split into individual currency units and were then settled by passing this unit back and forth between counterparties, this would require only a single unit of liquidity and thus be almost equivalent to offsetting in its liquidity effectiveness.

We define a large payment, x, as being of greater value than some threshold, T. We split each x into n equal-sized pieces such that $\frac{x}{n} \leq T$ and $\frac{x}{n-1} > T$. (For instance, for a payment, x, of value £190 million, and a threshold, T, of £100 million, there would be $n=2$ equal-sized pieces, each for value of £95 million.)

We investigate two potential benefits of such an approach: (i) a reduction in queuing due to part payments settling and releasing liquidity which would have been 'idle' without splitting; and (ii) the reduction in liquidity requirements if large payments are split and released piece by piece over time.

1.2   An illustrative example

Bank A has to make a payment of 150 to Bank B. Bank B has to make a payment of 100 to Bank A. Bank A makes its payment and then Bank B responds by sending its. In this case Bank A has a liquidity requirement of 150 units.

If instead Bank A and B were both to split their payments in half, Bank A could send 75 units and Bank B could respond with a payment of 50. Bank A could then send the other 75 units and Bank B could send the final 50 unit payment. In this case the liquidity requirement, still shouldered by A, has dropped to 100 units.

Without splitting:

Marginal liquidity requirement

| BANK A | 150 → | BANK B | 150 |

| BANK A | ← 100 | BANK B | TOTAL: 150 |

With splitting:

| BANK A | 75 → | BANK B | 75 |

| BANK A | ← 50 | BANK B | |

| BANK A | 75 → | BANK B | 25 |

| BANK A | ← 50 | BANK B | TOTAL: 100 |

BANK OF ENGLAND
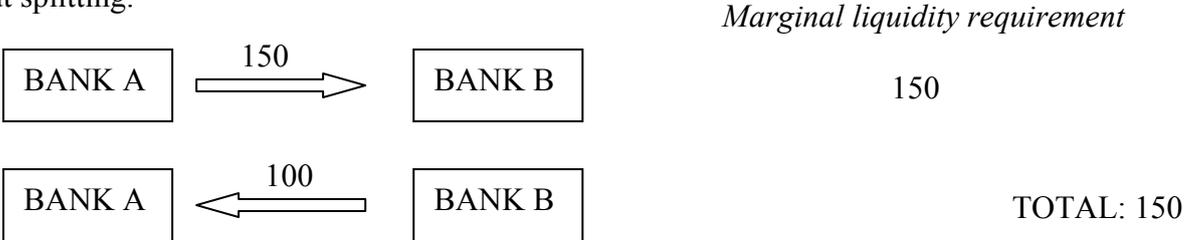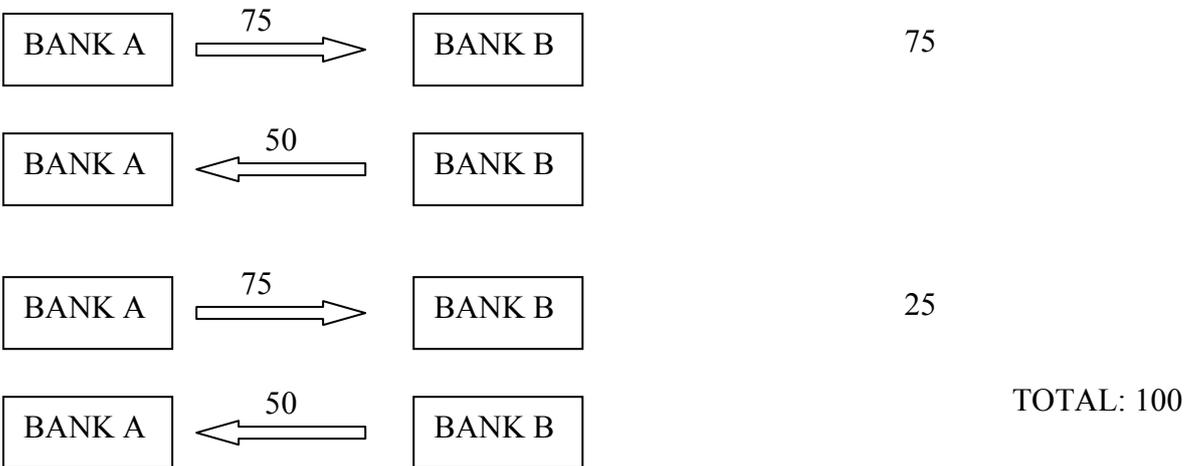
Although payment splitting is not a new idea in RTGS payment systems, it is not one that has been the focus of much research. Many studies such as Jackson and Ercevik (2009), Martin and McAndrews (2007), Johnson, McAndrews and Soramäki (2004) and Bank of Japan (2006) have considered the effect on payment systems of introducing other liquidity-saving mechanisms, but little attention has been given to payment splitting. The notable exceptions are Leinonen and Soramäki (1999) and Koponen and Soramäki (1998). Using the Bank of Finland Simulator, Leinonen and Soramäki (1999) show that there is considerable scope for payment splitting to help resolve gridlock situations and reduce settlement delay when a bank has insufficient liquidity. They show that splitting could eliminate payment queues at high and medium levels of liquidity. Koponen and Soramäki show that splitting the top 5% of payments by value can result in a 50%-60% reduction in average delays.

Section 1 of this paper outlines where payment splitting is currently used in systems worldwide; Section 2 outlines our methodology, summarising how CHAPS works, the generic liquidity requirements of different types of payment and settlement system, our data set and then the specific methods used to measure the reduction both in gridlock and then in liquidity requirements that can result from payment splitting; Section 3 presents our findings; Section 4 briefly examines legal issues relating to payment-splitting mechanisms; and finally Section 5 offers our conclusions.

2   Current usage of payment splitting

2.1   *Continuous Linked Settlement (CLS)*

CLS is a multi-currency cash settlement system. Using payment versus payment settlement it is able to eliminate much of the settlement risk that would otherwise be present in foreign exchange trades.

CLS receives a final list of the day's trades by 06:30 CET. It uses this to prepare a pay-in schedule for each participant that spreads the total liquidity requirement across five pay-in deadlines. For each currency a currency splitting threshold is set, in order to save liquidity (and reduce the scale of gridlock in the event of a pay-in failure). Any eligible transaction that is above this threshold value, in either currency, is split into smaller, equally sized transactions, each below the threshold value. These split transactions are spread randomly throughout the settlement queue. The CLS systems process the queue in order and settle transactions as and when there is sufficient liquidity to do so. CLS report that from January to May 2009 a little over 1% of trades are split on average – equivalent to approximately 2,500 trades a day.

2.2   *Swiss Interbank Clearing (SIC)*

Payment splitting also occurs in Switzerland's large-value payment system, SIC. It is not a formal process built into the system itself as in CLS, but rather a preference expressed by the Swiss National Bank. They request that, to reduce the probability and impact of gridlock, banks split payments above 100 million CHF (£59 million/€66 million) into smaller pieces. While this is not always adhered to, it is believed that such payments are generally split.

2.3   *BoJ-Net*

In Japan the requirement to split large payments has a similar level of formality to that in Switzerland. The Japanese Securities Dealers Association set out 'guidelines for ensuring smooth settlements' in its publication *The Japanese Government Securities Guidelines for Real Time Gross Settlement*. This sets an upper limit to the size of Japanese Government Bond (JGB) transactions that should be settled through the BoJ-Net payment system. Any transaction that exceeds ¥5 billion (£33 million/€37 million) should be split into smaller transactions and settled separately. This aims to curb the accumulation of outstanding settlements intraday by reducing the likelihood of gridlock. Again, this happens outside the system, at the level of individual banks.

2.4   *Canada ─ CDSX*

In late 2007 the Canadian securities settlement system CDSX amended the rules that govern the settlement of debt securities. The updated rules dictate that trades above CAD 50 million (£28 million/€32 million) must be split into blocks of CAD 50 million and a remaining 'tail' (if trades are not an integer multiple of CAD 50 million). This amendment aimed to reduce the risk of settlement failure and eliminate the need for participants to delay large-value trades until the end of the day. The Bank of Canada report anecdotal evidence suggesting that the introduction of splitting has been helpful for settlement efficiency in CDSX. Like in BoJ-Net and SIC this is not part of the functionality of the system itself and system participants are expected to split payments themselves.

2.5   *CHAPS*

There is no functionality in CHAPS to split payments, nor is there a rule or guideline recommending splitting above a certain threshold. Nevertheless anecdotal evidence suggests that some settlement banks occasionally seek agreement from their customers who are making large payments to split them to aid their intraday liquidity management.

To conclude, although the practice of payment splitting is not widespread, it does exist. Among systems, only CLS formalises payment splitting by providing the functionality to allow it to be performed in a controlled fashion. Among other things, such functionality facilitates reconciliation of the parts of a split payment. This in turn reduces legal and operational risks. The manual splitting practice undertaken by some CHAPS banks suggests there could be some (latent) demand for splitting functionality, to help reduce the liquidity needs in relation to large-value payments. We now consider what sorts of liquidity benefits payment splitting could bring.

3   Methodology

We run simulations to test two hypotheses:

   i)   Payment splitting can reduce queuing if there is insufficient liquidity available to settle all payments in a timely fashion; and

   ii)   If payments are split and then released piecemeal, then there will be a reduction in the liquidity required to settle all payments on time.

We do this using the Bank of Finland Payment and Settlement Simulator (BoF-PSS2). This simulator was developed by the Bank of Finland to enable central bank researchers and academics to investigate the effect on risk and liquidity needs of changes to banks' payment profiles or the specifications of the payment system itself.

A limitation of this simulation, and simulations of this type, is that we are unable to model behavioural changes. With payment splitting as an option, banks may be encouraged to send payments earlier in the day. In particular they may settle large payments piece by piece, as liquidity becomes available, rather than waiting to send the entire payment at once.

3.1   *The payment system*

CHAPS payments are submitted to the Bank of England to be settled subject to sufficient available liquidity. Most banks have their own internal scheduling systems that control the release of payments, although there is a central scheduler that is used by a number of banks. Banks are able to prioritise payments with codes 10-89 (with a low code signifying high priority). When payments queue, the following criteria are used to order the funds queue:

   i)   Priority code ascending (low code/high priority first);

   ii)   Amount ascending (smallest first); and

   iii) Submission time (earliest first).

By allowing smaller payments to be settled first if there is a queue and there are no prioritised payments, large payments do not block the movement of liquidity (although, in practice, to date, liquidity has been sufficient to allow even the largest CHAPS payments to settle in a timely fashion). These queuing criteria are applied to our simulations.

3.2   *Liquidity requirements*

At any point in the day the liquidity, L, required to settle all payments up to time, t, is a function of incoming and outgoing payments. Specifically,

$$L_t = \max{}_t (0, Out_i - In_i)$$

Where Out_t and In_t are the cumulative values of outgoing and incoming payments, respectively, up to time, t.

The liquidity requirements of a large-value payment system are bounded by two extremes. At one end is the 'RTGS requirement': the amount of liquidity needed in order that all payments are settled gross in real time, with no delays. This is equal to the largest net sender position for each bank during the course of the day. Mathematically:

$$L^{RTGS} = \max(L_t)$$

At the other end of the spectrum is the 'DNS requirement': the amount required to settle all payments by the end of the day.[5] This is equal to zero for banks which are net receivers by the end of the day, and to the end of day net sender position for banks which are net senders. Mathematically:

$$L^{DNS} = L_T$$

Where T is the time at the end of the day. By definition

$$L^{RTGS} \geq L^{DNS}$$

The following stylised example of net flows for a bank across the day shows this graphically.

Chart 1: A stylised example of a bank's intraday liquidity requirements
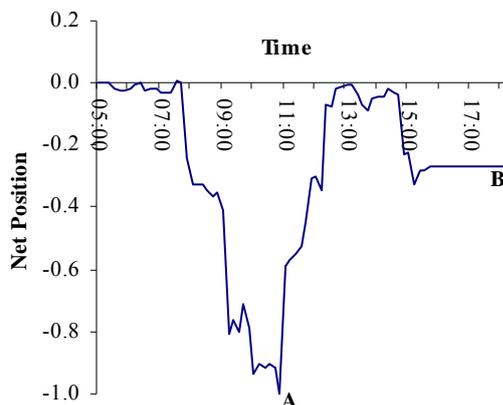
It shows that this bank had a maximum net sender position of 1.0 but a cumulative, end of day position of 0.26.

---

[5] For this to be entirely true there must be forced settlement at the end of the day. If not, in certain circumstances gridlock may occur, such that no bank has sufficient liquidity to make the first payment in the chain that releases liquidity for all the others. For example, if Bank A has to pay Bank B £100 and Bank B has to pay Bank A £100, then (despite the net flow being zero) one of the banks has to have £100 to allow the payments to be settled.

Here, **A** represents the 'RTGS requirement' and **B** the 'DNS requirement'. Payments will queue if the amount of liquidity available to a given bank is less than its 'RTGS requirement'. The values in the queue and the time that payments queue for will be a function of liquidity available and the timing and value of incoming and outgoing payments. If the amount of liquidity available is less than its 'DNS requirement' then some payments will be left unsettled at the end of the day.

3.3 Data

In the first part of the analysis, we have used payments data from four days spread across 2008-09. We have used data from the largest and smallest days by value in the year from June 2008 to May 2009 (30 June 2008 and 11 November 2008, respectively) and two days that were chosen randomly. These two days were near average in both volume and value terms (23 January 2009 and 29 April 2009). For the second part of the analysis, we expanded the data set to include an entire typical week in May 2009 (18-22). The data set that we used has all payments that were made by each participant in CHAPS on the given days. As well as the value, sender and recipient, we also had information about the time of settlement and the priority code.[6] We did not use data from more than four/nine days for the two parts of our analysis (respectively), because of the labour-intensive process of data-cleansing and running the simulations.

Transfers that were made from a bank's CHAPS settlement account to its CREST[7] account were not included in the data set. The liquidity required to finance such transfers would be dependent on payment flows in CREST. Analysis of such flows is beyond the scope of this paper.

While this data set has the time that each payment was settled, most banks use internal schedulers. We therefore have no information about whether payments were queued internally before they were settled in CHAPS. We also have no reliable information about whether the payments are made on behalf of customers or whether they are proprietary transfers.

---

[6] Priority codes do not appear to be actively used by CHAPS banks at present: the majority of payments have the system's default priority setting of 50.

[7] CREST is the United Kingdom's securities settlement system, providing a Delivery versus Payment (DvP) settlement service for gilts, equities and money market instruments. Banks that are direct participants in both CHAPS and CREST are able to transfer liquidity intraday from their CHAPS account to their CREST account and *vice versa*.

3.4   Hypothesis I: reduction in queues

If there is insufficient liquidity available to settle all payments as they are submitted, then payment queues will form. Payment splitting should help to reduce queuing, since splitting a payment into smaller pieces may allow partial settlement. Not only does this reduce the length of the queue directly, but it also frees liquidity that could be used to settle a payment in the recipient's queue that might otherwise have queued.

Currently there is sufficient liquidity in CHAPS to make payment queues a rarity: most banks actively schedule their payments through the day (using sophisticated internal schedulers), thereby ensuring that the ebb of the payments they send is sufficiently co-ordinated with the flow of payments they receive in, to keep their liquidity needs down while meeting all their payment obligations in a timely fashion. By encouraging banks to submit their payment instructions earlier in the day, liquidity-saving mechanisms (such as offsetting algorithms) would lead to queues of payments forming centrally (where they could be matched up and settled) rather than in the internal schedulers of the banks themselves (where there is no scope to achieve an offset). This is desirable, as it creates the potential for greater co-ordination of payment flows, leading to lower liquidity needs, without necessarily leading to payments being settled later than would be the case in a 'plain vanilla' RTGS system. Very large-value payments may struggle to find offsetting payments. If these are split, they can either be partially settled using available liquidity or offset with smaller incoming payments. Splitting payments may therefore also allow payments that are in the central queue to settle more quickly and with less liquidity.

We tested this hypothesis using a similar methodology to Koponen and Soramäki (1998). Bounding our simulation by the two extreme liquidity requirements described above (the upper bound of the 'RTGS requirement' and the lower bound, 'DNS requirement'), we vary the liquidity available, L^A, to each bank in the simulation between the lower and upper bounds in ten equal steps, ie

$$L^A = L^{DNS} + x(L^{RTGS} - L^{DNS})$$
$$x = 0, 0.1, 0.2 ... 0.9$$

Lower values of liquidity cause greater payment queues as there is insufficient liquidity to settle all payments as they are submitted. Each successive increase in liquidity results in a reduction in queuing. We do not allow $x=1$ in our simulations as by definition this results in no queuing.

We aim to show that at various levels of constrained liquidity, payment splitting will reduce the resultant payment queues. Table 1 shows the upper and lower bounds of liquidity requirements for all banks' on each of the four days simulated.

Table 1: Total amount of liquidity required by all banks to settle (£mn)

	30 June 2008	11 Nov. 2008	23 Jan. 2009	29 Apr. 2009
RTGS requirement	26,138	19,445	21,811	15,661
DNS requirement	12,779	2,417	7,605	4,628

Koponen and Soramäki (1998) use a settlement delay measure to evaluate the length of queues. We use a slightly different measure that captures both the value and duration of payment queues, the concept of the 'time value of queues'. This was used by Bedford, Millard and Yang (2004) to quantify the impact of an operational outage on large-value payment systems. It is equal to the value of a queued payment multiplied by the length of time that the payment queues for. Consequently a £1 million payment that queues for 10 minutes is considered to have the same impact as a £10 million payment that queues for 1 minute. Mathematically this is expressed as:

$$Q = \sum_{t=0}^{T} V_t$$

Where Q is the time value of the queue, V_t is the value of the queue at time t and the difference between t and $t+1$ is equal to 1 minute. The resulting value has the unit '£ mins'.

We simulate a base case whereby the time value of the queue is calculated at each liquidity level between the upper and lower bounds in the absence of payment splitting. This is used as a benchmark against which we evaluate the effect of payment splitting.

We split all payments that are above a given value into smaller pieces. They are then scheduled to be settled at the same time. If there is sufficient liquidity then all the parts of the split payment will be settled simultaneously, as though they were not split. If the bank is liquidity constrained at the point of settlement then a part of the payment may settle where otherwise none would have. This should result in a reduction in the time value of the queue. We split payments above three thresholds to investigate the effect on payments queues: £500 million, £250 million and £100 million. We chose these values as a significant number of payments were above them, and

thus eligible for splitting, yet the largest payments were not split into too many pieces. We should note that even the lowest of these values is well above the splitting thresholds that are in place in systems that already do payment splitting. Table 2 summarises the number and value of payments that were split on each day that we simulated.

Table 2: Volume and value of payments split at each payment threshold

	30 June 2008		11 Nov. 2008		23 Jan. 2009		29 Apr. 2009	
	Volume	Value (£bn)	Volume	Value (£bn)	Volume	Value (£bn)	Volume	Value (£bn)
Total	204,571	416	90,377	155	150,869	281	128,302	207
Above £100m	888	202	342	79	595	154	455	109
Above £250m	225	98	101	43	185	90	140	60
Above £500m	44	34	25	17	54	43	32	22

We assume that the Bank of England, as a member of CHAPS, always has sufficient liquidity to make all its payments and thus its payments were never split. As CLS Bank only ever pays out what it has received in first, it is never a net sender, and so splitting would have no impact on their liquidity requirement; we therefore did not split their payments either. We also assumed that payments to CLS Bank would not be split as they have to be settled by a strict deadline; however, payments to the Bank of England could be, and were, split in our simulation.

To measure the variation of the time value of queues across days we report the coefficient of variance of the data set. This is the ratio of the standard deviation to the mean and as such is dimensionless and scale invariant. This allows us to compare the variation of results at the different splitting thresholds.

$$C_v = \frac{\sigma}{\mu}$$

where C_v is the coefficient of variance, σ is the standard deviation across days and μ is the mean.

We average this across all liquidity levels for a given splitting threshold to see if one threshold gives more consistent queue reductions than the others.

3.5   Hypothesis II: reduction in liquidity needs

The liquidity needs of an RTGS system are defined as the sum of the each bank's largest net sender position. This need is often driven by the timing of the largest payments relative to incoming payments. If the largest payments are made before receiving substantial inflows then the liquidity requirements will be larger than if the incoming payments arrive first.

We know from anecdotal evidence that banks know the full details of many payments at the beginning of the day. Most banks tend to release small-value payments first thing in the morning but hold on to the larger, more liquidity-intensive, payments until after the CLS pay-ins and other time-critical transfers (such as clearings) have been made – normally by 9.30am UK time. This also allows them to wait for some incoming liquidity to fund these large payments. From a financial stability point of view, a delay in making payments is undesirable as it increases the impact of operational risk,[8] and can make liquidity management by banks more volatile (cf, Angelini (1998)). If a payment-splitting device were to exist, banks may be encouraged to submit parts of their larger payments earlier in the day. Their counterparties might then use that liquidity to make part of large payments that they were otherwise intending to make later, thereby creating a virtuous circle of settlement. If two banks both had large payments to make to each other, then in the absence of offsetting mechanisms, they may be able to pass split payments between each other until both payments are settled. This could result in substantial liquidity savings.

In the second part of our simulation, we expand on the work of Koponen and Soramäki (1998) and consider the liquidity impact of this 'split and spread' strategy. Firstly, we split all payments above £100 million into equal-sized pieces as before, each less than the threshold. As in the previous method (and for the same reasons), payments to and from CLS are not split, neither are payments made by the Bank of England. The first part of the payment is submitted on time and the remaining parts are submitted one every x minutes until the entire payment has been submitted. We simulated the release of remaining payment parts at 3, 5 and 7-minute intervals. This is an entirely artificial construct, yet one that aims to illustrate how the process of staggering the submission of large, split payments can result in liquidity reductions. This comes as a result of the arrival of incoming liquidity between the submissions of pieces of the split

---

[8] ie, if there were an operational outage during the day that prevented payments from being settled that day, then the impact of that outage would be greater, the greater the value of payments outstanding. For further discussion of the timing of payments see Becher *et al* (2008).

payment. The partial settlement could stimulate a payment cascade as illustrated in Beyeler *et al* (2007).

For example, as illustrated in Table 3, Bank A wants to send a £400 million payment at noon. Payments are split above £100 million and are spread at 3-minute intervals.

Table 3: The scheduling of a split £400 million payment submitted at noon

Time	Value
12:00	£100 million
12:03	£100 million
12:06	£100 million
12:09	£100 million

In this case final settlement has been delayed by 9 minutes.

In practice, what may happen is that a bank sends a partial payment, and if the counterparty has one or more large-value payments to make back then it uses this incoming liquidity to fund them. This liquidity can be passed back and forth rapidly until both payments are settled.

In the event that a payment was part delayed until after 16.20, when CHAPS routinely closes, the simulation was set up such that remaining parts of the payment all settled at 16.20.

Splitting using this alternative simulation approach clearly introduces some delay into the system as payments are spread over time. This delay is defined as being the time from settlement of the first piece to settlement of the final piece of the split payment. The average delay is relatively small: 8 minutes 50 seconds for payments split and spread every 3 minutes; 14 minutes 43 seconds when spread every 5 minutes; and 20 minutes 36 seconds when spread every 7 minutes.

We allowed all banks to have unlimited liquidity to ensure that they were never liquidity constrained and thus there were no queues. We then measured the required value of the liquidity 'RTGS requirement', as outlined above, by recording the largest net sender's position for each bank in turn. The system-wide requirement is simply the sum of the individual banks' 'RTGS requirements'. We exclude the liquidity requirement of the Bank of England and CLS Bank (the

latter of which is zero anyway) from this analysis. We then compare the system liquidity requirements with and without payment splitting to assess if there is any liquidity saving and, if so, quantify it.

CHAPS' throughput guidelines require that all settlement banks have made at least 50% of their payments by value by noon each day and 75% by 2.30pm. If, over the course a month, a bank did not meet these guidelines they would be expected to explain their reasons for not doing so. We look at the impact that delaying payments to save liquidity has upon the time that the 50% throughput guidelines are met.

4   Results

We present results at a system-wide level and, for the liquidity-saving simulations, also at a bank level.
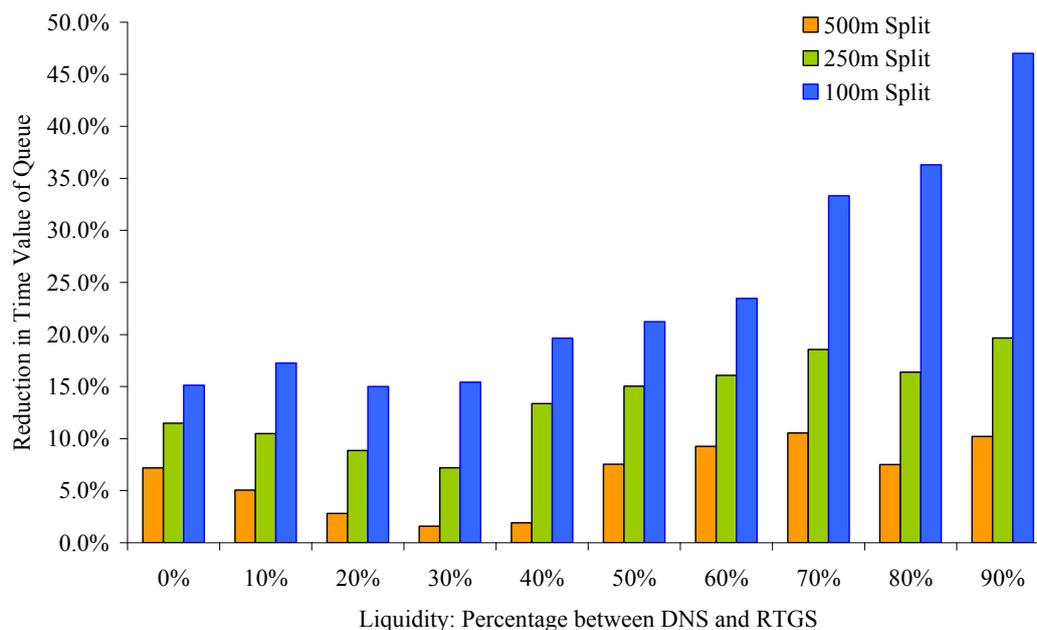
4.1   Hypothesis I: reduction in queues

Charts 2 and 3 below show the simple average of time value reductions in queues across the four days that we simulated.

Chart 2: Mean time value of queues at various liquidity levels, with and without payment splitting
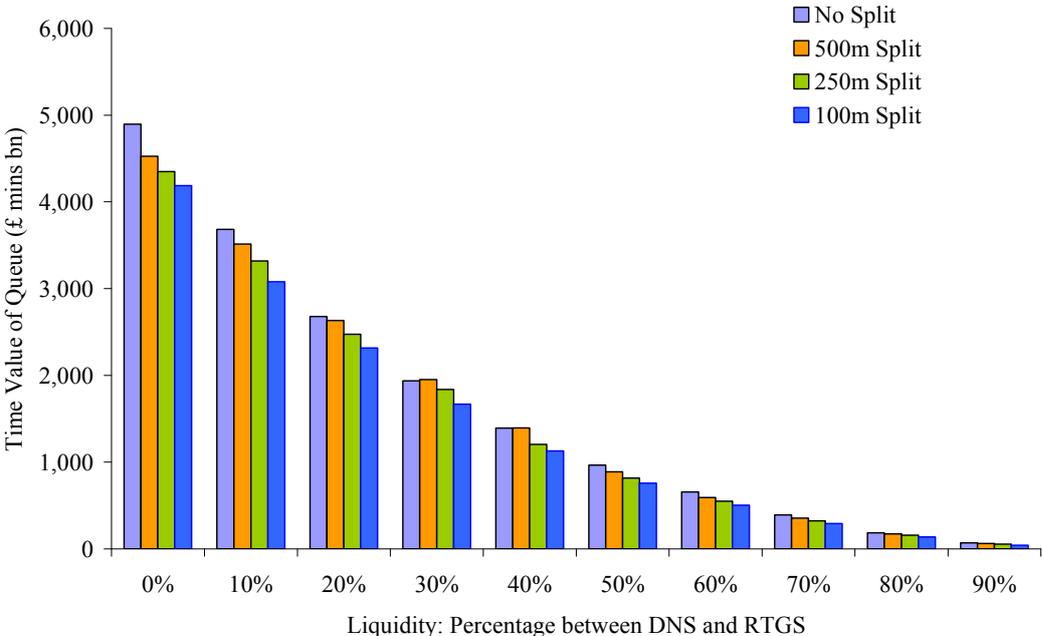
Chart 3: Mean reduction in the time value of queues due to payment splitting

As Charts 2 and 3 show, payment splitting has significant potential to decrease the duration and value of payments that are queued. For instance,

General observations:

- The time value of queues increases as the amount of liquidity available to each bank is reduced. This occurs non-linearly, accelerating as liquidity approaches the lower bound. Whereas the percentage reduction in the time value due to splitting increases at higher liquidity levels, in absolute terms the reverse is the case.

- There is a reasonably consistent pattern to the proportion by which payment queues are reduced under the different splitting thresholds across the various liquidity levels. In general the percentage savings are more pronounced at high and low levels of liquidity and less so at intermediate levels (20%-30%). The reasons behind this are unclear and may be due to idiosyncrasies of the small data set that we used; Koponen and Soramäki (1998) found a qualitatively similar trend using artificially generated payments data, and attributed this explanation to the observed pattern.

- When payments are split at the £500 million level there are modest reductions. In percentage terms these tend to be slightly larger at higher levels of liquidity. This is due to much shorter

queues, where the settlement of a single partial payment can have a greater proportional impact on the overall queue length.

- As the threshold level drops to £250 million and £100 million, not only are more payments split, but also those that are split are split into smaller, more granular pieces. These two factors explain the increase in the reduction of the time value of queues that we observe.[9]

Quantitative observations:
- If banks put only the minimum DNS amount of liquidity into the system this creates significant queues. The total time value of the queue is equivalent to all payments on an average day queuing for 17 minutes or £10 billion of payments queuing for over 8 hours. This can be reduced to about 14 minutes or 7 hours respectively if payments above £100 million are split.

- At the 70% liquidity level, the queues are rather more modest. They are equivalent to all payments queuing for 1½ minutes or £10 billion of payments queuing for 39 minutes. Payment splitting above £100 million can reduce this to 1 minute or 29 minutes respectively.

- Queues are reduced on average by 47% when there is 90% liquidity and payments above £100 million are split. In the majority of cases there was at least a 10% reduction in the time value of manifested queues in the simulations.

Koponen and Soramäki's 1998 paper split at much lower thresholds than our simulations resulting in an even greater reduction in payment queues. This is consistent with our findings that lowering the threshold results in an increased impact of splitting.

These average charts hide daily variations. Although the general trends across days are the same, the range of savings across days varies. On the smallest-value day (11 November 2008), the introduction of payment splitting at £500 million and low liquidity levels would in fact have resulted in an increase in queues, whereas on one of the near average value days (29 April 2009) the savings would have been above average. Splitting payments at the £100 million level leads

---

[9] Increased granularity means that there is greater potential to minimise the amount of idle liquidity that is held by banks when gridlock forms. For example if a bank has £200 liquidity and a £500 payment to make, splitting it into two payments of £250 will not release any liquidity. However, splitting it into five £100 payments will allow two partial payments to settle and thus the idle liquidity to circulate.

to the most significant, and consistent savings. The table below shows the coefficient of variation of queues arising from split payments for each liquidity level and splitting threshold. If the standard deviation is high compared to the average value then it means that there is more volatility across days; if this ratio is lower then the series is less volatile.

Table 4: Coefficient of variation of the time value of queues across days

Liquidity[a]	£500 million split	£250 million split	£100 million split
0%	0.80	0.67	0.47
10%	1.88	0.78	0.51
20%	2.61	0.94	0.69
30%	6.29	1.31	0.48
40%	5.41	0.26	0.29
50%	0.68	0.35	0.23
60%	0.63	0.46	0.31
70%	1.00	0.58	0.55
80%	0.61	0.27	0.62
90%	0.60	0.22	0.49
Average	**2.05**	**0.58**	**0.46**

[a]Liquidity refers to the increase in liquidity above the DNS level. 0% refers to the DNS level and 100% would refer to the RTGS level.

This shows that splitting payments above £100 million yields relatively more consistent queue reductions compared to splitting at the other levels.

Overall these results suggest that there is considerable potential for payment splitting to help reduce liquidity requirements without materially delaying the settlement of payments. It is not, however, a panacea. It is possible, in certain circumstances, for payment splitting to introduce queues, although this is rare (the annex provides a stylised example). The degree of reduction in queues increases as the threshold for splitting decreases and thus the number of payments eligible for splitting increases and the size of the consequent pieces is smaller. Compared to the average savings, splitting at a lower threshold results in less volatile savings and so is a more reliable approach. This is an intuitive result as the lower the threshold, the less liquidity there will be sitting idle on a bank's account. This leads us to conclude that in principle the optimal splitting threshold may be a single unit. However, as mentioned in the introduction in practice there would be a number of potentially material costs associated with such a threshold. A thorough investigation of these costs is beyond the scope of this paper.

4.2   Hypothesis II: reduction in liquidity needs

Chart 4 shows liquidity reductions on each of the nine days that we simulated.

Chart 4: Reduction in liquidity needs of CHAPS banks as a result of payment splitting
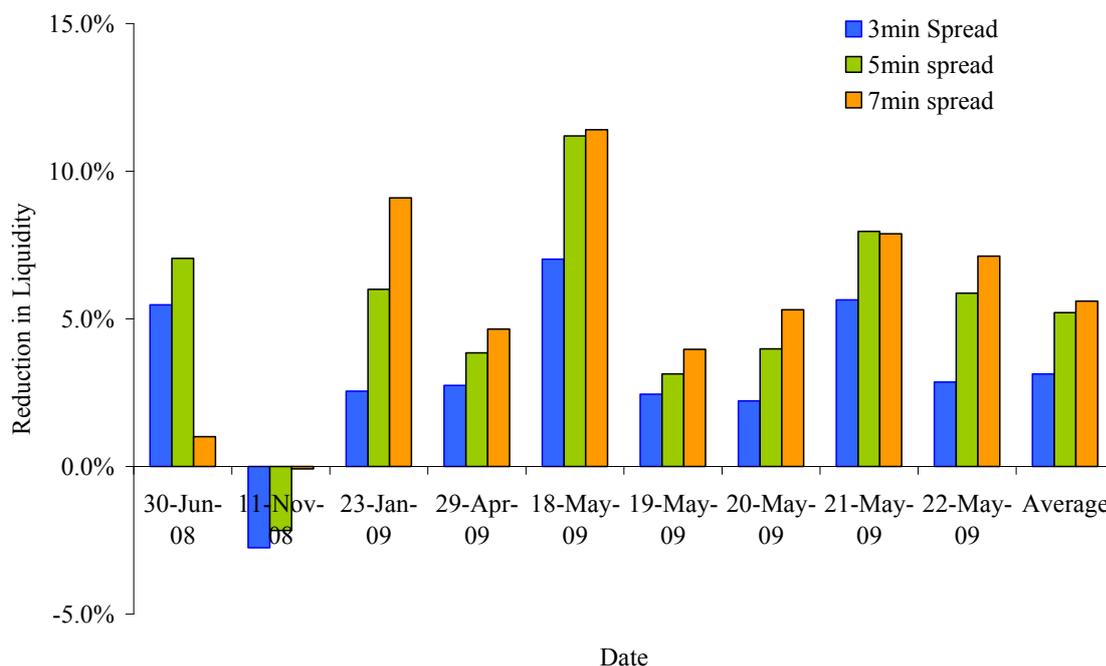

Chart 4 shows that there are modest, but not insignificant, reductions in liquidity that can be obtained by splitting payments. On average CHAPS requires over 5%, or £1.1 billion, less liquidity to settle all payments. On 21 May, all payments could have been settled with over £2 billion less liquidity than without payment splitting.

With the exception of 11 November 2008, which was the smallest day by value in the sample period from which the days in this analysis were drawn, every day that we simulated experienced a net saving in liquidity as a consequence of splitting payments. The general trend was that increasing the time interval between the settlement of each part of the split payment resulted in an increase in the liquidity saved, although the marginal increase declined as the time interval grew. This was due to the impact on the recipient bank: while most banks can absorb small delays in incoming payments without increasing their liquidity requirements substantially, as that delay increases the cost to the recipients starts to outweigh the benefit to the sender. On two of the nine days, spreading payments every 7 minutes instead of every 5 minutes resulted in a drop in the system-wide liquidity saving.

The benefits of payment splitting are not evenly spread across banks; there is a considerable degree of heterogeneity. While the majority of banks in our simulations benefit from splitting, some banks appear to be consistent losers. Banks that rely heavily on incoming payments to

fund their outgoing payments during the day are more vulnerable to a delay in incoming payments. If the simulation allowed them to change their behaviour (for instance, by delaying their outgoing payments in response) their liquidity requirements could remain unchanged. We might imagine that in reality such banks, which appear to be adopting a receipt reactive payment strategy, would adjust their behaviour to limit the impact of delayed incoming payments. Chart 5 shows that, while many banks stand to benefit (in some cases substantially) from payment splitting, this may be partly at the expense of others.

Chart 5: Average individual bank reduction in liquidity requirements due to payment splitting (selected CHAPS banks) – simulation results
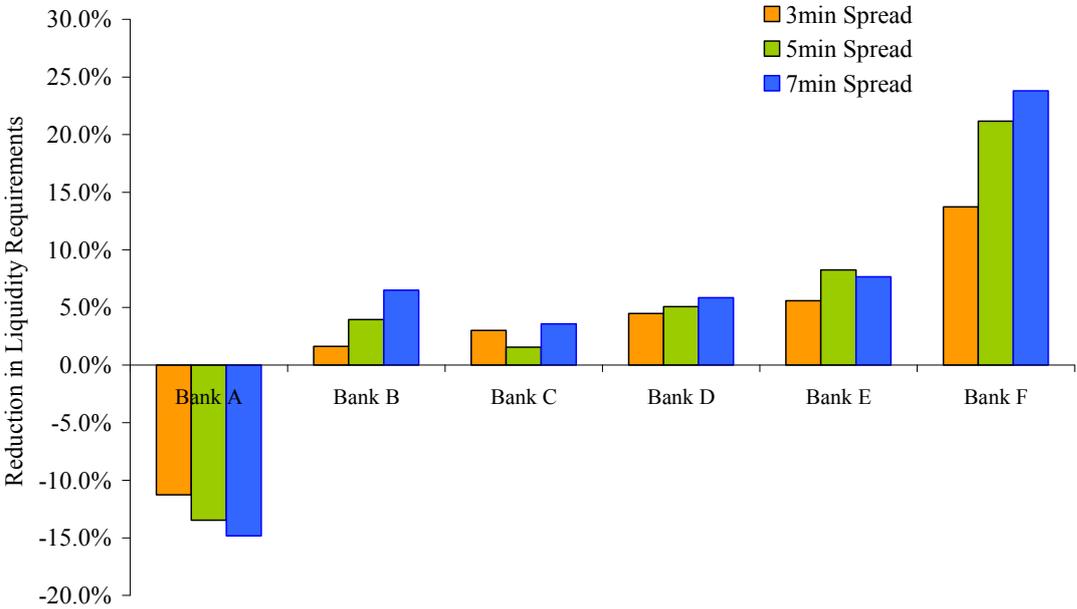
There was also considerable heterogeneity across different days. Chart 6 shows the maximum and minimum liquidity savings, in value terms, across the nine days simulated.

Chart 6: Maximum and minimum liquidity savings by bank (selected CHAPS banks) – simulation results
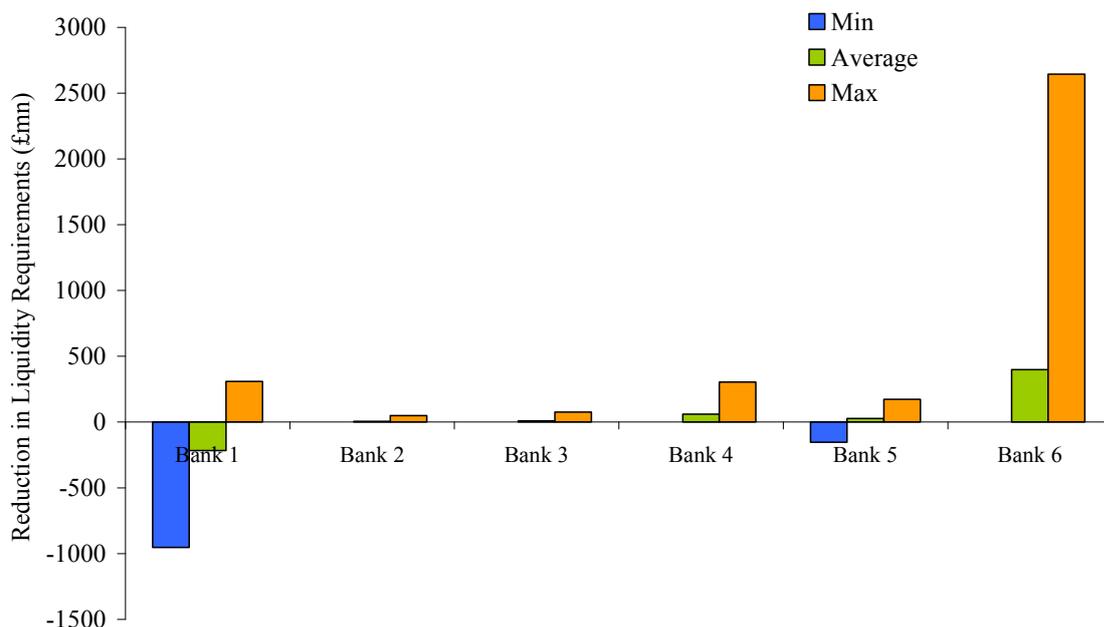
The majority of banks never suffered any increase in liquidity and in the worst case experienced no change in their liquidity requirements. And those banks that were most at risk of having their liquidity requirements increase also had the potential to achieve liquidity savings.

There is an inherent trade-off between settlement delay and liquidity saving. At the extreme all payments are delayed until the end of the day and the liquidity requirement is the minimum 'DNS requirement'. We have shown that payment splitting combined with the introduction of some delays can reduce liquidity needs. We would expect to observe the impact of the delays on overall settlement time in CHAPS payment throughput.

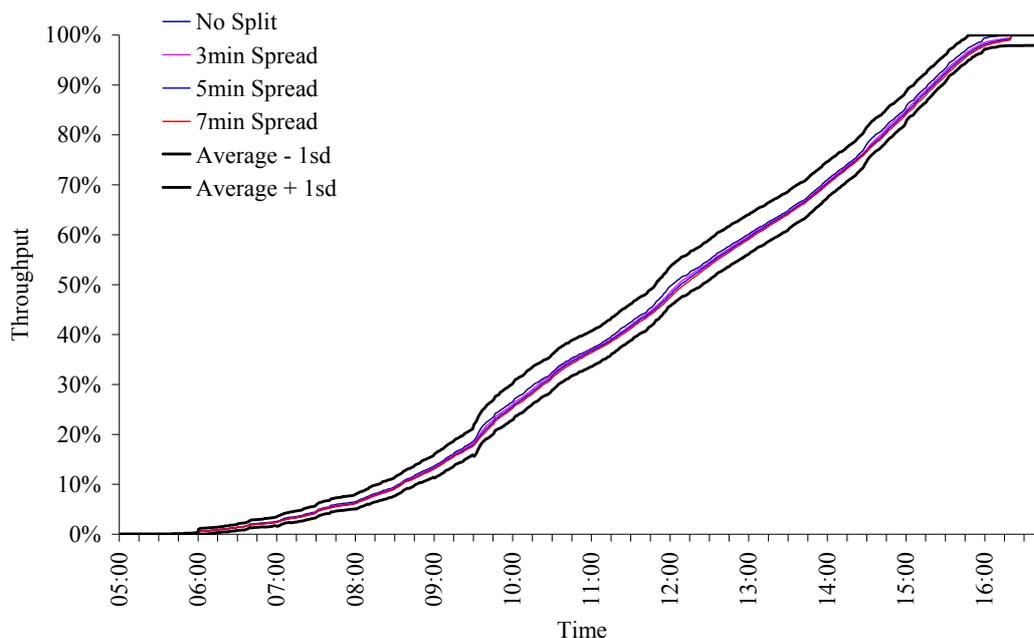**Chart 7: Average throughput with and without splitting**

Chart 7 shows the change in throughput due to payment splitting. By construction, the introduction of delays, as split payments are released piece by piece, results in a slight shift in average throughput. However, this effect is minimal. The black lines show one standard deviation of the normal variation in CHAPS payment throughput from 1 January 2008 to 13 May 2009. The delays introduced by this simulation are within a single standard deviation (and consistently by some margin) compared to the normal day-to-day variation in the timing of CHAPS banks' payments business. On average during the sample period chosen, 50% throughput was achieved at 12:06. One standard deviation difference in throughput comes 19 minutes either side of this. This compares to a 5-minute delay in the achievement of 50% throughput if payments are split and spread at 7-minute intervals. In short, payment splitting of this kind introduces delays that are well within normal variation of banks' payments business.

5   Legal issues and settlement risk

Payment splitting can present potential legal issues. If a payment is split, so that a portion of it is settled and then either the sending or receiving bank goes bankrupt prior to complete settlement, the question then arises; '*What is the legal status of the payment?*'. With its financial stability objective, a central bank would want to ensure that any legal risks arising from payment splitting would be appropriately managed. The degree to which legal issues (and hence risks) arise from payment splitting depends on the type of transaction that underlies the payment that is settled. What follows is a description of the legal risks which could arise. It does not represent a definitive judgement as to whether such risks could be fully mitigated.

There are two types of settlement risk relevant to payment splitting: principal risk and replacement cost risk. Principal risk is the risk that one party loses up to the full value of the trade if it satisfies its obligation but its counterparty does not. Replacement cost risk is the risk that the failure of a counterparty forces a bank to execute a replacement trade with a different counterparty, possible on less favourable terms.[10] The introduction of real-time and delivery versus payment settlement has all but eliminated principal risk from modern large-value payment and settlement systems. However, if a counterparty fails to fulfil obligations, then replacement cost risk remains.

Conceptually, payments fall broadly into three splitting categories: those where the underlying transaction is divisible, those where the underlying transaction is not divisible and those where there is no underlying transaction. Table 5 summarises the risk impact of payment splitting on each transaction type.

---

[10] See Bank of England (2005) for further explanation.

Table 5: Risks related to splitting different types of transactions

Underlying transaction	Example[a]	Risk implications of partial settlement	
		Principal risk	*Replacement cost risk*
Divisible (ie underlying transaction can be split).	Securities purchase, foreign exchange trade.	None, provided that supporting legal documentation is tight.	If payment splitting encourages earlier settlement, then replacement cost risk could be reduced.
Indivisible (ie underlying transaction cannot be split).	Business purchase, commercial property purchase.	Would create principal risk (since partial payment would need to be returned).	No implications.
None (ie no underlying transaction to split).	Repayment of unsecured loan.	If payment splitting encourages earlier settlement, then principal risk could be reduced.	None.

[a] While these examples serve to illustrate our distinction between transaction types, we acknowledge that the 'splittability' of the underlying transaction may not be as straightforward as presented. For example a securities dealer may not be prepared to make a trade whereby there was a chance, however small, that only part of that deal would be settled.

In the first category, where the underlying transaction is divisible, for example the purchase of securities, then the legal documentation underpinning the split payment could also require the securities transfer to be split (in the event of default) to reflect the split in the payment. In this case the principal amount would not be at risk, though there would remain replacement cost risk for the remainder of the outstanding transaction (though less than would have been the case if, under no splitting, the transaction had not settled at all at the time of default). CLS transactions are an example of this. As CLS either settles both sides of related foreign exchange transactions or neither, there is never any principal risk. If a CLS participant becomes insolvent having settled part of a split transaction and not the remainder, its counterparty is exposed to the market risk in replacing the remainder of the trade. CLS ensures that bilateral arrangements exist between the two counterparties of a particular split currency transaction. Specifically, they expect that the issue of partial settlement is written into the ISDA or IFEMA[11] agreements that bilaterally provide the terms and conditions that dictate the transaction. In SIC the decision to

---

[11] International Swaps and Derivatives Association (ISDA) agreements and International Foreign Exchange Master Agreements (IFEMA) are standardised contracts that detail the terms and conditions of the trade. This includes the consequences of default or *force majeure*.

split is left at a bank level. This means that responsibility for ensuring that all parties are clear about what should happen in the event of partial completion of payment is left to the banks themselves.

In the second case, a payment has an underlying transaction that is not suitable for splitting. This is where the issue of the legal status of a partial payment is particularly pertinent. Each 'partial payment' may in itself be recognised as a unique transfer order and thus (in the EU) be subject to EU Directive 98/26/EC on settlement finality in payment and securities settlement systems (known as the Settlement Finality Directive). This ensures that, once final, payments are irrevocable. However, where a partial payment does not fully discharge underlying contractual obligations, this creates legal uncertainty. In turn, the recipient might either ring-fence those funds that it has received until the whole payment has settled, thereby not benefiting from the liquidity saving; or use those funds, while exposing itself to the risk that the sender goes into default and the part payment is revoked, thereby forgoing the credit risk advantages of an RTGS system. To avoid these risks, system rules might need to be introduced that prohibited the use of payment splitting in relation to payments whose underlying transaction are indivisible. Were a system to consider implementing payment splitting, further thought would need to be given to whether this was feasible, and whether it could fully mitigate the legal risks.

The third category of payments are those which do not have an underlying transaction, for example the repayment of an unsecured overnight loan. In CHAPS, these payments are typically for very large values (cf, Millard and Polenghi (2004)), the payment order is known at the beginning of the day and, although there is a market convention that repayment should occur the following morning, the time of settlement is, in practice, relatively flexible. In this case, in as much as payment splitting may encourage early submission of payments, there could even be a reduction in credit risk as a result of payment splitting. Specifically, if a bank makes a partial repayment of funds earlier than it would have done without payment splitting there is a reduction in the credit exposure between the lender and borrower. In the event of default, the lender has received partial repayment of the loan which is a better outcome than joining the unsecured creditors for the full amount. However, before any system adopted payment splitting, proper consideration would need to be given to the legal status of the unpaid portion.

6 Conclusions

This paper does not attempt to make the case for payment splitting to be introduced into any particular payment system. Rather, it contributes to the growing literature on mechanisms for making RTGS payment systems more liquidity efficient, highlighting the potential benefits along with areas that would require further consideration.

Payment splitting has the potential to reduce the liquidity requirements of an RTGS system. The lower the threshold for splitting, the greater the potential benefit, as payment splitting not only directly reduces any payment queue when a part payment is settled but also, in the process, releases idle liquidity which may be used to settle further payments sooner than would otherwise be the case. When combined with an offsetting algorithm applied, say, to a non-urgent payments stream (in which queues could be expected to be commonplace), payment splitting may help to ensure that payments are offset more efficiently, further reducing liquidity needs and the amount of time that large payments queue. By making offsetting more efficient, payment splitting is likely to lead to liquidity savings over and above those of a stand-alone offsetting algorithm. This is because the most difficult payments to offset are likely to be the largest-value payments, those which are most suitable for splitting. This would need to be demonstrated via an extension of this study.

The results presented in this paper suggest there is no Pareto improvement in liquidity requirements. That is, while most banks benefit, some banks may not. This result may reflect the static assumptions of bank behaviour that this type of simulation employs. We acknowledge that our methodology is an artificial construct to see what the effect of payment splitting could be in the absence of other LSMs and behavioural changes. It is entirely possible that a bank which receives a part payment would change its behaviour to send a part payment sooner in return. This would result in a further reduction in liquidity usage.

Although payment splitting is practised in a few systems, notably CLS, there remains the question of the legal status of partial payments. Bilateral agreements between counterparties and between settlement banks and their customers could provide a solution; so too might system changes to control any residual risk that payment splitting creates. Where payment splitting encourages earlier settlement of payments, then for some payments, such as the repayment of unsecured interbank loans, such earlier settlement would serve to reduce the credit risk that banks are exposed to.

Furthermore, while we have anecdotal evidence that some banks in the United Kingdom already split payments internally to aid their liquidity management, it has been noted by others that there are concerns about the reconciliation of parts of a split payment that make splitting unattractive. This could be remedied by a centralised system that tracks and 'reassembles' split payments.

There is a careful balance to be struck between the system-wide benefits and cost to individual banks; between timely settlement and liquidity saving; and between benefit to the sender and cost to the receiver of the delay in sending/receiving the full value of a payment. In addition, there are some payment types that are not suitable for splitting from a legal/risk perspective. But structured correctly, and subject to a more detailed consideration of the question of legal risk, payment splitting could in principle reduce the overall liquidity required to settle payments in real time, with an insignificant effect on the timing of settlement. Indeed, once behavioural changes were allowed for, payment splitting may lead to earlier settlement of large-value payments resulting in a reduction of operational risk and facilitating banks' liquidity management. Given that, in practice, payment splitting would probably best be applied in conjunction with another liquidity-saving mechanism, such as an offsetting algorithm, an extension to this study could be to look at the extent of liquidity savings when the two approaches are combined.

Annex   How can payment splitting result in increased queuing?

While rare, it is not impossible that payment splitting causes payment queues to *increase*. In our simulations, at the 30% liquidity level splitting payments above £500 million resulted in such an increase. The following stylised example illustrates how this can happen:

Bank A has to make two payments. The first is for a value of 150 units and is submitted for settlement in period 1; the second is 100 units and is submitted for settlement in period 2. It starts the day with 120 units of liquidity. In period 6 the bank receives a payment of 130 and is able to settle all its payments.

Without payment splitting:

	Liquidity	Payment due	Settled?	Queue
Period 1	120	150	No	150
Period 2	120	100	Yes	150
		150	No	
Period 3	20	150	No	150
Period 4	20	150	No	150
Period 5	20	150	No	150
Period 6	150	150	Yes	0

In this case the 150 unit payment cannot be settled until period 6. This results in a total time value of the queue of **750** units.

With payment splitting (payments over 100 units are split into equal-sized pieces each less than or equal to 100 units):

	Liquidity	Payment due	Settled?	Queue
Period 1	120	75	Yes	75
		75	No	
Period 2	45	100	No	175
		75	No	
Period 3	45	100	No	175
		75	No	
Period 4	45	100	No	175
		75	No	
Period 5	45	100	No	175
		75	No	
Period 6	175	100	Yes	0
		75	Yes	

In this case the 75 unit part payment and the 100 unit payment cannot be settled until period 6. This results in a total time value of the queue of **775** units. Payment splitting has, in this case, increased queuing.

References

Angelini, P (1998), 'An analysis of competitive externalities in gross settlement systems', *Journal of Banking and Finance,* Vol. 22, Issue 1.

Bank for International Settlements (2005), 'New developments in large-value payment systems', *Committee on Payment and Settlement Systems, Publications No. 67.*

Bank of England (2005), *Payment Systems Oversight Report 2004*, January.

Bank of England (2009), *Payment Systems Oversight Report 2008*, April.

Bank of Japan (2006), *Japan's Next-Generation RTGS,* October.

Becher, C, Galbiati, M and Tudela, M (2008), 'The timing and funding of CHAPS Sterling payments', *Federal Reserve Bank of New York Economic Policy Review,* September*, pages 113-33.

Bedford, P, Millard, S and Yang, J (2004), 'Assessing operational risk in CHAPS Sterling: a simulation approach', *Bank of England Financial Stability Review,* June*, pages 135-43.

Beyeler, W, Glass, R, Bech, M and Soramäki, K (2007), 'Congestion and cascades in payments systems', *Physica A: Statistical Mechanics and Its Applications,* Vol. 384, No. 2, October.

Jackson, J and Ercevik, K (2009), 'Simulating the impact of hybrid design on the efficiency of large-value payment systems', in Leinonen, H (ed), *Simulation analyses and stress testing of payment networks: proceedings from the Bank of Finland Payment and Settlement System Seminars 2007-2008,* pages 189-228.

Japanese Dealers Association (2004), *The Japanese Government Securities Guidelines for Real Time Gross Settlement*, January.

Johnson, K, McAndrews, J and Soramäki, K (2004), 'Economising on liquidity with deferred settlement mechanisms', *Federal Reserve Bank of New York Economic Policy Review, December,* pages 51-72.

Koponen, R and Soramäki, K (1998), 'Intraday liquidity needs in a modern interbank payment system: a simulation approach', *Bank of Finland Studies in Economics and Finance E:14.*

Leinonen, H and Soramäki, K (1999), 'Optimising liquidity usage and settlement speed in payment systems', *Bank of Finland Working Paper No. 16/99,* November.

Martin, A and McAndrews, J (2007), 'Liquidity-saving mechanisms', *Federal Reserve Bank of New York Staff Report No. 282.*

Millard, S and Polenghi, M (2004), 'The relationship between the overnight interbank unsecured loan market and the CHAPS Sterling system', *Bank of England Quarterly Bulletin,* Spring, pages 42-47.